Utilizing web analytics: a case study of a digital construction permit web service

Tuomas Paukkeri
University of Tampere
Faculty of Communication Sciences
Human-Technology Interaction
M.Sc. thesis
Supervisor: Jaakko Hakulinen
June 2017

University of Tampere
Faculty of Communication Sciences
Human-Technology Interaction
Tuomas Paukkeri: Utilizing web analytics: a case study of a digital construction permit web service
M.Sc. thesis, 67 pages, 5 appendix pages
June 2017

This thesis inspects the usefulness of web analytics data in a context of an eGoverment web service. The aim of the study was to find out what can be learned about the user behavior and how to utilize the captured user interaction data when considering future improvements to the service to enhance the overall user experience.

The research of the thesis is based on a case-study inspecting the web analytics data of the Lupapiste web service. The Lupapiste web service allows citizens to apply for constriction permits electronically, while municipalities and cities are able to handle construction permits and the communication by utilizing the Lupapiste web service. Google Analytics was used as the main source of the user interaction data.

The results showed web analytics data can give a detailed view of the user behavior, depending on the details of the web analytics implementation and coverage. Inspecting the web analytics data from multiple viewpoints and from varying levels of detail, the data can create insights about the user behavior. The insights from the data can be used for guiding the future improvements of the service. Furthermore, the segmenting of the web analytics data enables to discover distinctive user behaviors. Also, the data can be utilized when considering more detailed inspection of the user behavior such as usability studies. However, the web analytics data cannot reveal the motivations of the users, thus web analytics data should be used in a combination with other research methods to capture attitudinal data of the user behavior.

Key words and terms: web analytics, users behavior, data analysis, user experience

Contents

1. Introduction 1
2. Strategic approach to web analytics 5
    2.1. User experience and user engagement 5
    2.2. How to approach web analytics strategically? 7
    2.3. Why to use the web analytic data? 9
    2.4. Data driven design, data informed design 10
    2.5. Key performance indicators (KPI) 12
    2.6. Metrics and vanity metrics 12
3. Web analytics in practice 14
    3.1. Trace data and unobtrusive data collection 14
    3.2. Instrumentation of web analytics 15
    3.3. Web log analysis and web usage mining 15
    3.4. Page tagging 17
    3.5. Tag managers 19
4. Case study: Analyzing user experience of Lupapiste 21
    4.1. Lupapiste service 21
    4.2. Research questions and methods 23
    4.3. Instrumentation of web analytics 25
5. Results 27
    5.1. The number of users and sessions across a timeline 27
    5.2. Returning versus new users 30
    5.3. Usefulness of segmenting the web analytics data 30
    5.4. Usage hours during the days of week 32
    5.5. Usage patterns based on event tracking 34
    5.6. Event flow from analytics 41
    5.7. Device usage 42
    5.8. How users arrive to the service 45
    5.9. Engagement based on bounce rate 46
    5.10. Engagement based on visit duration 47
6. Discussion 50
7. Conclusions 54

References 56

Appendices

1. Introduction

Web data analytics have an important role in the modern-day web development process. Let us pretend that we have a website, a catalogue full of specific goods to be sold online which visitors browse and possibly purchase items from the website. From the website owner's perspective, orders are being handled, money exchanges owner and orders are shipped to customers. However, successful purchases made by the visitors are not the only thing that can be learned about the website. How many visitors are actively browsing but not completing the purchase process? In what phase of the ordering process do users abandon the shopping cart? What product categories or products are viewed by the visitors the most? What other content on the site are the visitors browsing?

By implementing and utilizing web analytics to the website described above, the owner of the website can gain detailed information about the visitors of the site and the features relating to the purchasing of the products. When the interactions are being captured, the site owner can inspect and evaluate how users actually end up with a successful purchase. That opens the possibilities for further improvements and enhancements based on the real user interactions.

Web analytics refers to the automatic data collection of interactions between a system and the users of such a system. Web Analytics Association defines web analytics as "the measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing Web usage" (Association, 2008, p. 3).

The popularity of web analytics is increasing both in the usage of web analytics and companies providing web analytics tools. New web analytic services seem to be popping up in the Internet frequently. The estimated value of web analytics market will be close to 3 billion euros in 2019 (MarketsAndMarkets, 2016).

In order to capture meaningful and actionable data with the use of web analytics, a strategic approach to linking the business goals of the web service to the implementation of the web analytics can be a helpful way to also include different stakeholders to the process. By utilizing the stakeholders, it is possible to create mutual understanding about the success of the website based on the web analytics data.

In an environment where the developed web services vary greatly from a simple marketing websites containing only a few selected pages to large-scale information systems, the use of analytics cannot be overlooked especially from the viewpoint of, for example, evaluating the successfulness of design decisions and the service itself. In an optimal situation, the use of web analytics should be considered already in the design phase in terms of what makes a feature or interaction successful and which features enable users to reach the designated goal in the service. It is essential to consider what should be measured in order to gain such an insight that allows us to evaluate the successfulness.

As designers plan the interactions and developers implement the features and interactions, evaluating design decisions and the features themselves can enhance the overall user experience. Analyzing the gathered data might also help guide the further development efforts of any given features or interaction of an information system. Providing the development team information on how users are using the newly implemented feature, is essential when evaluating if the new feature is improvement to the previous version or not.

The added value of the web analytics is that the captured user interaction data can provide meaningful insights about user behavior and increase the understanding about the users (Jansen, 2009, p. 2). Web analytics can also reveal possibly problematic areas or features from the website or web service. Problems might occur when users are not able to finish the desired tasks or meet the desired goals for which the site has been built for. In addition, web analytics can also be used for researching purposes. Technical issues, such as errors, might be affecting the use or users are having hard time finding relevant content from the website. Web analytics can also give directions and hints for what to inspect in more in-depth guiding the possible development areas of the web site. Furthermore, small-scale usability study findings can be examined and verified through use of web analytics data. (Cardello, 2013)

Web analytics allows inspecting the data and behaviors in time perspective. Developers and designers can inspect events in temporal manner, events occurring daily, weekly, monthly or even yearly. Custom time ranges can be also used when filtering and reporting interactions and behaviors (Lalmas, O'Brien, & Yom-Tov, 2014).

As noted by Singal et al. (2014), the web analytics tools and services are improving continuously and providing more sophisticated analysis of the user behavior and site performance. However, the tools and services won't be beneficial if the business goals of the site aren't fully understood and transformed into measurable metrics. Based on the measurable metrics, it is possible to gain the biggest value in the form of insights.

The big deficit of web analytics is that it cannot give answers to the *why* questions. Why did the users behave the way they did? Why didn't users complete the task they started? Web analytics cannot offer any information about the motivations of the users or shed any light on the decision process of the users. In order to explore the motivations of the users, attitudinal data is required, such as surveys, interviews or direct in-context observations. However, the web analytics data can be the starting point for the gathering of the attitudinal data. Unclear behaviors discovered through the web analytics data can be the basis for attitudinal data gathering and research. (Jansen, 2009)

Currently there is a vast line of products available for capturing and analyzing the interactions and behavior of users in any given website. One particularly popular web analytics service is the Google Analytics (GA) service provided by Google. With the use of Google Analytics administrators, developers, designers or other stakeholders of a website can gain various insights about the usage of the website or web service. Google Analytics will be used throughout this thesis in the analysis and extraction of user behavior data.

However, despite the possibilities offered by web analytics, little research has been done on their use in an eGoverment context to create better services for citizens and authorities. This thesis explores what can be learned from the behavior of the users in modern web applications utilizing web analytics in the eGovernment context. In the context of this thesis, the term eGoverment refers to the electronic communication between the citizen and the government. (Wikipedia, 2017)

The case study in this thesis focuses on an eGoverment service called Lupapiste. Lupapiste web service is essentially replacing the citizens’ visits to government offices enabling the communications and interactions to happen solely through the use of a digital web service. Lupapiste web service enables citizens to apply for a construction permit when starting to build ones’ house. In addition to the citizens - designers and architects play a central role in the building process and are also important stakeholders in the application process. As a result, all the documentation, communications and interactions between citizens and officials take place solely in the Lupapiste web service.

In the ever-changing world of the modern-day web service creation, the development process is iterative. Further enhancements and improvements is a natural way to create better and more engaging experiences for the users of the service. Decision upon which features and interactions will be developed or perhaps removed from the service can be validated through the use of web analytics.

From the viewpoint of a user experience designer answering questions as to what is the regular path users use in the service are simple but highly interesting. Furthermore, the captured user behavior data provide an insightful picture of the usage of the service which designers and developers can utilize when considering the direction of upcoming development directions.

The following quote by Seth Godin is highly relevant in the context of this thesis. “Don't measure anything unless the data helps you make a better decision or change your actions” (Godin, 2014). When considering the web analytics, the aim is to capture data which can be used to make better decisions and essentially better service for the end users and customers. The scope of the analytics and what to measure is highly dependent on the service in question. Considering two different types of web services e.g. personal blog versus e-commerce site; what to measure with web analytics might totally differ between the websites.

The contents of this thesis is as following: Section 2 investigates two central terms, namely user experience and user engagement relating to the web analytics and inspects strategic approach to web analytics and why to bother with the web analytics. Section 3 focuses in the intricacies of web analytics also presenting web log analysis and web usage mining. Section 4 dives in to the case study inspecting the web analytics data from multiple viewpoints. Section 5 presents the overall findings from the case study. Section 6 includes discussion based on the case-study and findings. Conclusions will be drawn in the section 7.

2. Strategic approach to web analytics

In this chapter I will inspect the general approaches, concepts and principles when considering the utilization of web analytics in an any given web service, website or product.

2.1. User experience and user engagement

The term *user experience* relates to humans using an interactive product, a web service in the context of this thesis, and the emerging experience that comes through the use of interactive product. For that matter, the product can be any given device, website, application, software, any apparatus that the user is interacting with. The basic notion of user experience design is that without considering the user experience and possible emerging experiences, the device or product cannot mediate experiences. The user experience is temporal phenomena relating to the usage of an interactive product or systems with all the possible feelings, thoughts, motives and actions. (Hassenzahl, 2010)

User engagement on the other hand is the result of a shaped user experience. It can be argued that a shaped user experience will result in a desired user engagement. The user engagement can be seen as the quality of user experience that highlights the positive aspects of the interactions. The concept of user engagement relies on the observations that successful technology is not only used but engaged with. The user engagement as well as the user experience are affected by the context of use, the value of the system and the user interface of the system, for example. (Attfield & Kazai, 2011)

In the context of this thesis, the user experience will be ultimately shaped through various details through-out the process of the construction application in which the user engagement can vary depending on the current situation and context of the users and the state of the application.

In order to measure user engagement, a profound definition is needed. Lalmas et al. (Lalmas et al., 2014) dictates a clear definition of the term user engagement:

> “User engagement is the emotional, cognitive, and behavioral experience of a user with a technological resource that exists, at any point in time and over time.”

The definition considers the direct interaction between the user and the system as well as the desire of the user to return to the service.

User engagement is a double-edged sword: in some cases, longer user engagement does not necessarily mean that the user is truly engaged with the service, but simply frustrated with bad usability, for example, not finding the needed functionalities to complete the task at hand. In this sense, the user engagement is dynamic and situated in the given context. (Lalmas et al., 2014)

Mapping the user engagement to the context of web analytics, the following Table 1 presents web analytic metrics measuring the user engagement according to Lehmann et al. (2012). By measuring the metrics listed below, it is possible to create a model of user engagement as noted by Lehmann et al. (2012).

Metrics	**Description**
Popularity	(for a given time frame)
#Users	Number of distinct users.
#Visits	Number of visits.
#Clicks	Number of clicks (page views).
Activity	
ClickDepth	Average number of page views per visit.
DwellTimeA	Average time per visit (dwell time).
Loyalty	(for a given time frame)
ActiveDays	Number of days a user visited the site.
ReturnRate	Number of times a user visited the site.
DwellTimeL	Average time a user spend on the site.

Table 1: Engagement metrics defined by Lehmann et al. (2012, p. 4)

The implications of user engagement are highlighted by Warnock and Lalmas as following: “engaged users are (1) likely to view more content and spend more time on a website and (2) more likely to return to the website at a later date (Warnock & Lalmas, 2015). Warnock and Lalmas (2015) also note the difficulty of measuring the user engagement which is normally done with eye tracking studies, although the mouse cursor has been reported to correlate with eye movements. Also, physiological measurements such as facial expressions, muscle activity and self-report measurements are not in the scope of this thesis and therefore not discussed any further. (Lehmann et al., 2012; Warnock & Lalmas, 2015)

As this thesis utilizes Google Analytics, which does not provide features for the mouse cursor tracking, the measurement of the user engagement solely relies on the metrics provided by GA.

The user engagement concerning Lupapiste is difficult to conceptualize based on the description above. The service is not designed for viewing more content or spending more time in the service. The opposite might be the truth as users only view the required content and want to spent the least amount of time as possible. The returning of the users is in a way natural, as users are required to return to the service in order complete the required steps of the construction application or add the required attachments.

2.2. How to approach web analytics strategically?

Utilization of web analytics can be divided into two separate processes: before the launch of a new service or feature and exploring existing website or feature in an existing website. Although, this thesis will concentrate on the latter as the research explores an existing publicly available web service, it is important to acknowledge how to approach web analytics before publish the service to public.

With the use of strategic approach to web analytics, the services should be considered in the sense that what is useful to measure from the users' interactions with the service. Consider a website selling any kind of goods online, the most import thing capture from the web analytics might be how many of the users add items to a shopping cart and successfully complete the checkout process. What might be visible in the analytics data is that most of the users drop out from the checkout in a specific step. The designers of the service could then review the checkout process based on the analytics data in order to discover how to possibly improve the checkout process in order to decrease the dropout rate of the checkout process.

Before the above scenario would play out, the utilization of the web analytics should be planned out with a great detail and consideration. From a general perspective, the use of web analytics should be linked to the business goals of the service. Essentially linking the web analytics data to the reasons why the service or product exists at first hand. Modern day web analytics software enables the capture of vast amount of data which in its all entirety might not be useful. It is likely that only a portion of the available data gives the needed insights. The upper-management level personnel might be interested in a more general while, a user experience designer would be interested in a more detailed information about the successfulness of website, for example, on a feature level of detail. (Waisberg & Kaushik, 2009)

The Figure 1 below by Waisberg and Kaushik (2009) exemplify the approach of mapping the business goals to web analytics. It important to acknowledge that the

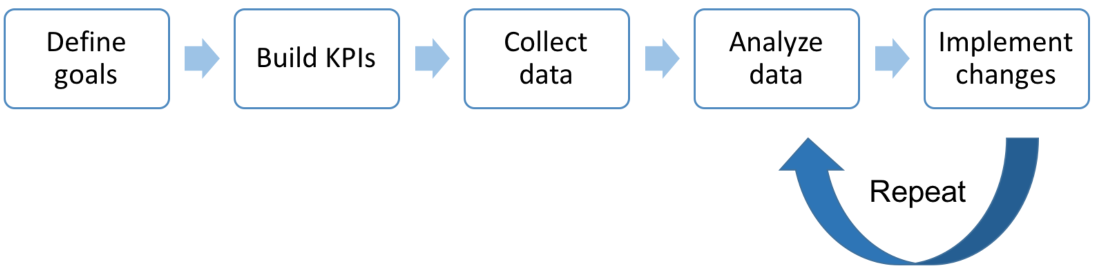


Figure 1: Web analytics process by Waisberg and Kaushik (2009)

process is not run-through just once, but in a constant cycle. This allows developers to react whenever the implemented changes take effect in better or worse or inspect whether the metric is useful at all. As Waisberg and Kaushik (2009) note, site owners should find the critical numbers which to follow through the web analytics.

Web analytics process can be concretized through a web measurement plan which enables to create focus for the web analytics and allows the stakeholders to concentrate on the metrics which are the most useful. Collecting all possible data may end up in a sea of data which is too difficult to analyze and extract meaningful insights. Web measurement plan essentially maps the *key performance indicators* to the web analytics through metrics. (Cardello, 2013)

In the case of Lupapiste the web measurement plan would enable the development team to focus on the most important and insightful metrics while ignoring the distracting metrics. By mapping the business goals of the service to the key performance indicators, the web measurement plan would also create transparency inside the development team and the company's internal stakeholders such as the management, marketing and customer support. Web measurement plan creation would also enable follow-through based on the metrics.

As Nakatani and Chuang (2011) noted the use of web analytics can be central method in creating competitive advantage in the rapidly evolving world. Carefully selecting the web analytics tools is vital as the decisions made in the selecting the tools can have long term influences for the business. From this viewpoint, it is also important to consider what web analytics service or software will be adopted. Nakatani and Chuang (2011) raise the issue of using a free web analytics service, which might not be effective decision depending on the business adopting web analytics. It is also important to consider whether the analytics software is provided as software as a service or installed in house. The latter might require additional resources in order to set up. Nakatani and Chuang (2011) also note the issue of the context web analytics in terms of the website's abilities contrasting the different web analytic requirements between retailer's website and manufacturer's website. (Nakatani & Chuang, 2011)

Even though the web analytics adoption might seem like straightforward process, it is important to consider the context of the adoption, what website or web service will be the target system as it might affect the tool selection for example. If the website is one of the critical components of the business, it should be evaluated whether a commercial web analytics solution would ensure competitive advantage in addition to informing decision making of the business.

Although, building the web measurement plan would create a structured approach to utilizing web analytics, Beasley (2013) introduces an alternative approach to web analytics shown in Figure 2.

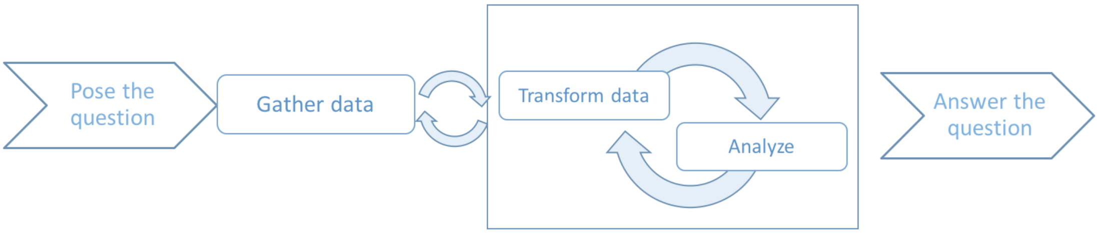


Figure 2: Web analytics process by Beasley (Beasley, 2013)

The methods enable explorative research to be performed by framing the context as a question. The question represents a gap in the current knowledge, essentially describing what you want to learn. The question to be answered will steer the data gathering to the required direction such as what metrics to use in order to find the answer. Depending on the situation, transforming the data might be a simple case of looking at an already available report or exporting and manipulating the data in a spreadsheet program. As noted by Beasley (2013, p. 16) the analysis stage might require you to transform and manipulate the data again. This stage is possibly the most time-consuming and difficult in the sense that when it is the right time to stop the analysis. Beasley (2013) suggests to tell a story with the data when answering the question. The format of the story depends on the question. Answering a simple question might be just reporting a single number while answering a complicated question might require more time for extracting the insights.

The method by Beasley (2013) show in Figure 2, was the central method utilized in the case-study of the thesis. The research was open-ended in the sense that no strict research plan was followed. Therefore, additional questions rising in the amidst of the research could be examined as well. However, primarily the research focused on the research questions from the section 4. Closing the gaps in the current knowledge of the user behavior in Lupapiste and finding out the current state of the user experience guided the research by utilizing the method shown in Figure 2.

2.3. Why to use the web analytic data?

The use of web analytics data main premise is to support decision making. However, it is important to highlight that web analytics is not the one and only user experience research method. Web analytics can be used to complement other user research methods and processes utilized such as surveys, usability studies, interviews and field observations (Beasley, 2013; Cardello, 2013; Kumar, 2010).

When inspecting web analytics data from a user experience designer's viewpoint, it can be used for evaluation and understanding, improving an experience and discovering something new or something not obvious, essentially creating a new perspective about people and users which are in the center of the design process. (Churchill, 2012; Pavliscak, 2014, 2015). Furthermore, web analytics can be used for inspecting whether the site is meeting the goals or is there something hindering users from reaching desired goals. In addition, web analytics data can be utilized in optimizing the site's features and overall user experience by validating hypothesizes. Insights from the web analytics can also be combined with other research methods as a source of inspiration, for example, what to study in more detail in usability studies. (Cardello, 2013; Pavliscak, 2014)

By utilizing web analytics data, it is possible to solve possible disputes with stakeholders. For example, when considering whether to remove a certain feature from a website, web analytics data can be used to evaluate if the removal is right or wrong choice. If users aren't using the particular feature at all, perhaps it is safe to remove. Web analytics also enables one to inspect whether the changes made to a service have made an improvement. When inspecting, for example, a sign-up process, did the changes lower the time it takes to complete the sign-up process. Web analytics can also be used to evaluate how the improvement should be made by comparing two alternative solutions by A/B split testing, in which the two solutions can be evaluated head-to-head with real users. Web analytics allow one to inspect relationships within the data and data points, discovering something previously unknown. This can be seen as the most complicated viewpoint to utilization web analytics. (Pavliscak, 2014)

Even though the support for the decision making is the primary use of analytics, web data analytics can provide new perspectives to a given problem or for inspecting a product from a different viewpoint.

2.4. Data driven design, data informed design

Utilizing web analytics data can also be seen as a way of doing and thinking in a more general level. The utilization of data can be expanded to cover all levels of the organization. This holistic approach is usually referred as a data driven design method in which the utilization of web analytics is a central factor in how design and the overall business is approached. The methods comprise tools and processes which enable steering the design, development and management through web analytics. The data driven design approach is particularly popular among start-up businesses which utilized the method in order to survive and find the right product for the right users (Croll & Yoskovitz, 2013).

Data driven design can be broken down into three step process: build, measure, learn, shown in

Figure 3. Having an idea which is built into a product which is then instrumented with analytics in order to learn how the idea works with real users. The cycle is run continuously enabling fast learning and also enables to evaluate whether pivot the idea to a different one. This cycle can also be used with an existing product or service, in which case the idea can be a hypothesis derived from a product or a website. Ideas based on the hypothesis will be tested and validated whether the hypothesis is true or false. (Croll & Yoskovitz, 2013; Hoverfält, 2015)

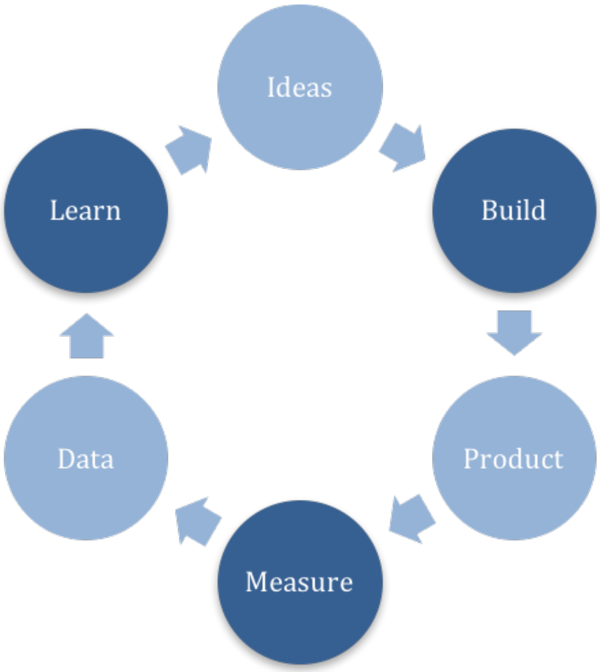


Figure 3: Build, measure, learn-cycle (Croll & Yoskovitz, 2013)

It is important to note that data driven design can lead into a situation where everything is based on data questioning the designers output on the design process. The famous example of this is Google testing 41 different shades of blue. The shades of blue with countless variations were tested to support decision making although the decision could have been done without. (Hern, 2014; Walker, 2009).

Data informed design method approaches the situation from a slightly different perspective. In the data informed design process, compared to the data driven design method, the data is seen as a single component in a big and messy problem being solved. Data informed design acknowledges that one has only a subset of information in order build successful product. The data is based on the currently built system and features of the system and the audience of the current system is behind the data. Hitting the local maximum through iterations is achievable but on a broader scale one can only have the current snapshot of the data. (Chen, 2012)

As Pavliscak conclude, the approach should not be either data or intuition but the combination of the two: data and intuition (Pavliscak, 2015). Designers utilize the intuition in creating design solutions for a particular problem. These solutions can then be validated with the use of analytics (Croll & Yoskovitz, 2013).

As noted by Hoverfält (2015), the data driven design can also be an organization wide approach to business. In the cross-section of design, business and analytics, exists the biggest value of the approach. Concentrating only on a single section of the previously mentioned, for example analytics, will not be the viable option as the biggest value will emerge with the combination of the three sections. In order to gain the most out from the data driven design, organizations have to have the will to change the organization itself to embrace data driven design.

In general, the utilization of data driven or data informed approach can be organization's way to doing business acknowledging the value of the data. From a more fine-grained perspective, it is essential for the designers to understand how to utilized data in combination with other method as well as follow-through the usage of the data.

2.5. Key performance indicators (KPI)

A lot of attention should be put into inception of *key performance indicators* (KPI) to find out sensible and actionable indicators for the current context of the business. KPIs should be derived from the business model of the service as noted in the Section 2.2. When the KPIs are linked to the business goals, the development of the service is steered towards the business goals. Through the definition of the KPIs, the definition of success is being dictated, this way the business can see whether the business is successful or not. According to Hurst (2010), the identification of the business metrics and key performance indicators is required. It is essential to inspect the overall business objectives, goals and business strategies and then evaluate how the user behavior enables achieving the overall business goals. (Hurst, 2010)

Waisberg and Kaushik (2009) divide a good KPI into four different attributes. KPI should be *un-complex* in the sense that all the stakeholders understand the meaning of the measurement and therefore can make decision based on the KPI. The KPI should be *relevant* in the context of the website and the business. On the surface two similar kind of websites will probably have different KPIs as they are linked to the business models of different businesses. *Timely* attribute relates to the KPI providing the needed information immediately in the sense that waiting for months for the information is already too late to make a decision. When a good KPI is *instantly useful*, anyone can grasp the essence and insight right at the first look. (Waisberg & Kaushik, 2009)

2.6. Metrics and vanity metrics

Metric term used through-out web analytics has been defined as a quantitative or numeric measurement of users' behavior expressed in numbers. Metric can be a single measurement, for example, the number of users or a ratio of new users versus returning users or a rate such as the bounce rate. (Beasley, 2013)

Essentially metrics should be comparative and understandable. One should be able to compare metrics between different time periods or user groups to help understand possible developments or changes in the measurement. It is essential that the metric is understandable in the way that different stakeholders can discuss the possible changes in the metric and change their behavior based on the metric. If no one understands the metric, it will not give the insight required to change things around. Metrics presented as a rate or ratio are insightful in the way that they are more comprehensible and allow one to act based on the value. As Croll and Yoskovitz (2013) conclude: “A good metric changes the way you behave”. When change occurs in an any given metric, what will be done differently based on the change of the metric is essential.

It is important to consider what metrics will tell correct information about the website, web service or product. Let’s imagine you are inspecting various metrics captured by the web analytics software. All the reported numbers have a lot of digits and one could presume this as a success. It is essential to ask what will be done differently based on the information. If one doesn’t have an answer, then the metric might not be useful. Vanity metric is a number which is not actionable and does not tell if something is good or bad. One cannot change anything based on the metric or number reported. (Croll & Yoskovitz, 2013)

The definition of metrics can be extended with leading and lagging indicators which are important to consider whether the metric inspected will indicate something about the future or reporting possible problem. Leading indicators or metrics provides signals which the company could act upon. Lagging indicators indicates possible problem with the website or product. For example, *churn* metric, which is used for reporting the number of users leaving a particular website. Problem with the *churn* for example is that users have already left the service leaving the developers or business without further touch points to the users. Nonetheless, lagging indicator might be the only available metrics in the early phase of utilizing web analytics. (Croll & Yoskovitz, 2013)

Whatever the KPIs and metrics are chosen it is important to note, that one have to have willingness to change the KPIs and metrics if the select indicators are not giving results or are measuring a wrong thing. While developing the KPIs and metrics, essentially goals are picked early in the process while the definition of success is not clear. Adjusting the KPIs and metrics should be done if wrong things are being measure. (Croll & Yoskovitz, 2013)

3. Web analytics in practice

In this chapter, the main methodologies and central concepts used in the realm of web analytics are presented. The theoretical background and fundamental elements of web analytics will be also discussed.

3.1. Trace data and unobtrusive data collection

Imagine a forest scenery with a fresh snow covering the ground. When you start walking through the forest, you are leaving footprints to the ground on top of the snow cover. The more people are walking the same track you walked, the more tracks are left to the ground creating traces of behavior. These physical footprints or the digital traces collected by the web analytics are traces of human behavior. In the context this thesis, these footprints on the snow, become data from which researchers can speculate and reason about the behavior of the users. Trace data is a collection of interactions between the user and web service. Users are leaving digital traces much like the footprints in the woods but in digital format. Contrary to the footprints in the woods metaphor where erosion creates the traces, in a digital context the trace data is accreting over time. (Jansen 2009)

The biggest advantage of collecting trace data is the unobtrusiveness, meaning researches can collect the trace data without a direct response or reactions from the participants, non-reactively from the perspective of the user. Contrasting unobtrusive method to an obtrusive method, for example a usability study conducted in a laboratory setting, the test participant is fully aware of being studied while the researcher is intruding the context of the user possibly affecting the participant's viewpoints and answers. The observer effect, phenomena where the researcher is in a direct contact with the participant, may affect the participants answers and actions during the study. (Jansen, 2009, p. 16) When utilizing unobtrusive data collection methods, data represents more natural behaviors of the users as opposed to laboratory setting. The data collection happens invisibly to the users on the background, thus avoiding the Hawthorne effect (Lalmas et al., 2014).

Web analytics enables researchers to capture data in the wild in the sense that web analytics does not need any special arrangement, except implementation of the web analytics service to capture the interaction data. A public web service, which is accessible for all and any user, the sample size can include all users of the service. The scale of online web analytics in which it can operate is limitless because all site visitor can be included in the data sample. Sample size can be big but this depends solely on the possible user base or popularity of the service. With small visitor amounts the optimizations number might not be big enough – the results may not be statistically significant. (Jansen, 2009)

3.2. Instrumentation of web analytics

Instrumentation of web analytics means applying methods and technologies for the purpose of recording and storing relevant data related to the interaction between the users and the system. Two different types of techniques utilized when capturing user behavior data or transactional data from an online service are log files and page tagging. Log files refer to the web server log files, which capture the user interactions, for example pages accessed by the users, in the form of a log file entry. In academia and literature, this method is known as transactional log analysis (Arshad & Ameen, 2015). The other common method used in the instrumentation is the page tagging method discussed in more detail in the section 3.4.

The advantages and disadvantages of the two methods are presented in Table 2 (Singal, Kohli, and Sharma 2014, p. 25).

Log files		**Page tagging**	
Advantages	**Disadvantages**	**Advantages**	**Disadvantages**
Requires no changes to the Website or extra hardware installation	Records interactions with the Web server	Near real-time reporting	Requires extra code to the Website
Requires no extra bandwidth	Server must be configured to assign cookies to visitors	Easier to record additional information	Uses extra bandwidth each time the page loads
Freedom to change tools with a relatively small amount of hassle	Only available to companies who run their own web servers	-	Hard to switch to Analytic tools
Logs both page request success and failure	Cannot log physical location	Able to capture visitor interactions within Flash animations	Can only record successful page loads, not failures

Table 2: Advantages and disadvantages of web log files and page tagging methods (Singal, Kohli, and Sharma 2014, p. 25)

3.3. Web log analysis and web usage mining

Web server access logs have been used to capture and analyze the usage of web sites. Web server log files capture every request that the server receives and log files collect users' interactions with the website. By default, the access logs contain a wide range of data. A single web log file entry can contain, for example, visitor IP address, date and time of the request, page requested and information about the user operating system and the browser. Depending on the configuration of the server software, four different types of log format are usually used: NCSA Common Log, NCSA Combined Log, NCSA Separate Log, and W3C Extended Log. The specific format of the log files can be customized by the server administrator to include more information about the requests. Figure 4 exemplifies a single web log file entry in the NCSA Combined Log format.

```
111.222.333.444 - - [12/Apr/2015:16:45:41 +0300] "GET /index.php HTTP/1.1" 200
570 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_2) AppleWebKit/537.36
(KHTML, like Gecko) Chrome/41.0.2272.89 Safari/537.36"
```

Figure 4: Example of a log file entry in NCSA Combined Log format.

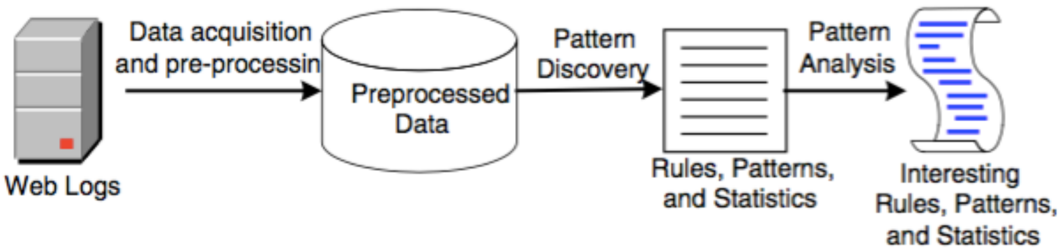


Figure 5: The Web usage mining process (Das & Turkoglu, 2009)

Basic level user behavior information from the web log files can be extracted by using tools such as AWStats (http://www.awstats.org/) and Piwik (http://piwik.org/). These tools generally report the number of visits, visit durations, visit times, most viewed pages, entry and exit pages, operating systems and browsers for example. In some situations, this level of fidelity can be sufficient for creating an overview of the user behavior.

However, log files capture every requested resource, which in turn will generate noisy data. In order to gain useful insight and information from the log file, additional data manipulation steps are required. By applying data mining techniques to the log file data, more detailed information can be extracted from the log files. This method is more generally referred to as web usage mining. The process of web usage mining is depicted in Figure 5.

Vellingiri et al (2015) divide log file processing into three steps: pre-processing, pattern discovery and pattern analysis. The discovered patterns are usually represented as collections of pages, objects, or resources that are frequently accessed by an individual user or, more frequently, by groups of users having the same needs or interests. (Castellano, Fanelli, & Torsello, 2013)

In the pre-processing phase, the log data is cleaned from unnecessary entries. These entries usually include multimedia files, images, stylesheets and requests with an irrelevant HTTP Status code and entries generated by the search bots. In the pattern discovery phase, the goal is to form meaningful clusters from the behavioral data and extract navigational patterns. The pattern discovery phase relies on applying an appropriate algorithm to the web log data such as Weighted Fuzzy Possibilistic C-Means Algorithm. By using algorithms in the pattern analysis phase, the user behavior can be predicted with quite accurate precision. Based on the request by the user, a list of helpful pages could be presented based on the pattern analysis essentially helping the user to navigate the web site. (Vellingiri et al., 2015)

The purpose of the pattern discovery phase is to cluster the user behavior and navigational patterns by means of other relevant fields such as statistics, machine learning, pattern mining and artificial intelligence. Vellingiri et al (2015) used Weighted Fuzzy Possibilistic C-Means Algorithm in their study. Clustering the user behaviors is an important step for the data analysis and insight creation based on the users' behavior. (Vellingiri et al., 2015)

Data recorded in log files may not be always entirely reliable. The problem of the unreliability of these sources of data is mainly due to the presence of various levels of caching both on the client side and possible proxy servers. When the request for a cached content is sent, the request might not reach the server at all when the cached version of the content is returned to the users. The misinterpretation of the users' IP addresses causes issues with log file data. Multiple users may use the same IP address in which case these entries may be counted as a single user. (Castellano et al., 2013) Overall, a web log in itself does not necessarily reflect a sequence of an individual user's documented access. Instead, it registers every retrieval action, but without a unique identification for each user. (Román, Dell, Velásquez, & Loyola, 2014)

3.4. Page tagging

One common way to capture the user behavior from a web site is the page tagging method. Page tagging relies on a piece of JavaScript that is inserted to the source code of the page. When a page is loaded by the web browser, the JavaScript code is executed by the browser and data is transferred to the server hosting the analytics application. Depending on the need of the web site administrator, the tracking can be targeted only to a part of the website. More commonly all the pages of a given web site are equipped with a tracking code. Applying page tagging to all of the pages accessed by the user essentially enables discovering the user behavior more holistically as all the possible pages are being captured. To capture the user interactions from every possible page, means that the administrators have to include the tagging script to all of the pages which might cause resistance because of the expenses of the work needed. (Jansen, 2009)

An example of a page tagging code, also known as code snippet, is presented in Figure 6 which exemplifies the default Google Analytics page tagging JavaScript code.

```
<!-- Google Analytics -->
<script>
(function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;i[r]=i[r]||function(){
(i[r].q=i[r].q||[]).push(arguments)},i[r].l=1*new Date();a=s.createElement(o),
m=s.getElementsByTagName(o)[0];a.async=1;a.src=g;m.parentNode.insertBefore(a,m)
})(window,document,'script','//www.google-analytics.com/analytics.js','ga');

ga('create', 'UA-XXXX-Y', 'auto');
ga('send', 'pageview');

</script>
<!-- End Google Analytics -->
```

Figure 6: Example of Google Analytics JavaScript (Google, 2016)

The code snippet works by dynamically inserting analytics.js file from the Google servers to the page. A tracker object is being generated by the snippet, which is used to send a hit to the vendor analytics service. The sent hit includes the user interactions of a given page triggered by the user behavior such as accessing the page or possibly more detailed record of the user behavior, for example downloading a file. (Google, 2016)

The popularity of Google Analytics is quite extensive in the top 10 million websites included in the W3Techs technology overview report. Majority of the websites (82.2%) are utilizing Google Analytics as the web analytics tool. (W3Techs, 2015)

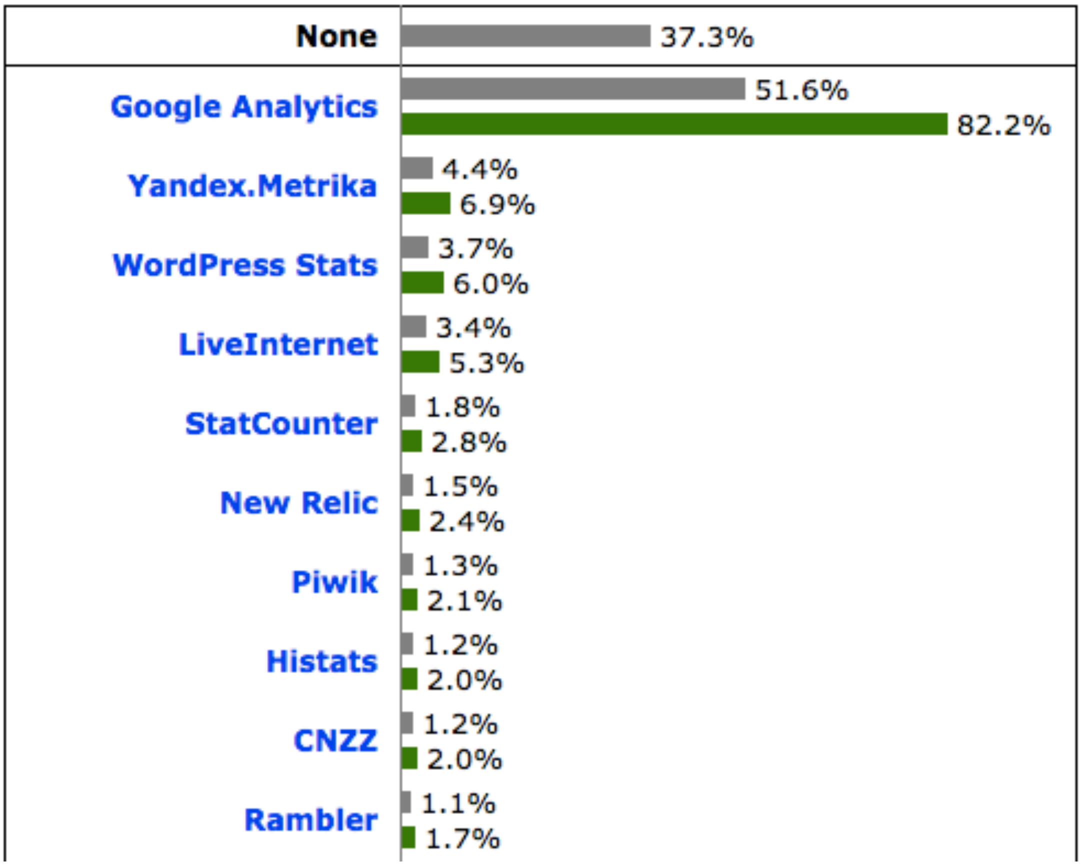


Figure 7: Top ten of analytics services in use by W3Techs. 51.6% of the websites are using Google analytics. Market share of Google Analytics is 82.2% (W3Techs, 2015)

3.5. Tag managers

Recently Google has been promoting the Google Tag Manager service as a way to implement users' behavioral tracking to a website. In addition to enabling web analytics implementation, Google Tag manager provides functionalities to include other tagging scripts, third-party scripts, to a website, third-party page tag scripts meaning other than Google services. Google Tag Manager is essentially user interface which provides tools and features for implementing more detailed behavioral tracking strategy. For example, one can track individual links or form submits and even capture user behavior based on clicks. Google Tag manager also provides functions for implementing different tag services through a single user interface. (Google, 2017c)

Google Tag Manager streamlines the page tagging method described in the section 3.4 by wrapping multiple tags in to a single container which is deployed to the website via a single JavaScript snippet. In a situation where website owner is using multiple tracking services, all of the tracking snippets would be included in the pages' source code adding the tracking snippets manually. When the Google Tag Manager (GTM) is used, the tracking codes are added to the container snippet, which handles the deployment to the website. As more tracking services are added to the website, the source code of the pages will stay the same and snippet updates are made through the

Google Tag Manager service. When deploying Google Analytics (GA) snippet through the Google Tag Manager container, the GA snippet has to be removed and replaced with the GTM snippet. It is important to note, that though GTM is used, tracking data is still collected and reported in the Google Analytics.

Furthermore, utilization of the GTM simplifies, for example the event tracking by enabling the creation of the event tags through the GTM administrative user interface. When adding new events to be captured in the website, the definitions of the events do not require code changes to the website. Adding new events does not require any programming skills or further technical experience essentially lowering the barrier of utilization of detailed event tracking. (Farney, 2016)

4. Case study: Analyzing user experience of Lupapiste

In this section, I will present the context of the research done in this thesis by describing the features and a brief history of the Lupapiste service. Research questions and the main research method are also presented.

4.1. Lupapiste service

The history of Lupapiste is linked with a Finnish governmental programme named the Action Programme on eServices and eDemocracy abbreviated as SADe. SADe program was directed by the Finnish Ministry of the Environment between 2009 and 2015. The Lupapiste was originally a pilot project for applying construction permits electronically. The pilot project was organized with 28 municipalities across Finland. After the governmental project came to an end, Solita Oy remained as the provider of the service and continued the management and marketing of the service. In the spring of 2017 Solita set up the Lupapiste as independent company now known as Evolta. Evolta is currently responsible for the development of the service with a staff of 17 personnel with on-going efforts to expand the coverage of the service nationwide. (Evolta Oy, 2017b)

The value proposition of Lupapiste service is that the service enables cities and municipalities to replace construction permit applications regularly handled with paper with an electronic alternative. With the use of Lupapiste the whole permit process of a construction project can be done electronically replacing the usual paper forms, construction drawings and all possible paper documents needed for applying construction permit. Evolta promotes the service as a virtual workspace for handling construction permits for citizens, companies and authorities. The interactive workspace enables citizens to ask guidance and support from the authorities related to the construction and the permit itself. The project space in the service can also be used for saving all required documents such as application files, required attachments and architectural plans. In addition, Lupapiste enables authorities to approve the applications in the service. The application process in Lupapiste is always up-to-date and enables transparent process for all the stakeholders related to the application and construction itself and also allows informal communications between the parties.

Lupapiste also enables electronic archiving of the construction permit documentation through the service. Cities and municipalities can discard the printing and archiving of the papers and attachments. (Evolta Oy, 2017b)

Lupapiste users can be roughly divided into two groups: applicant users and authority users. Applicant users are mainly citizens applying for construction permits. The construction permit can also be applied by architects or designers on behalf of the applicants. Authority users are government representatives who handle the official side of the permit process eventually presenting the verdict on the application. The applicant users can also include construction companies as Lupapiste enables the construction companies to handle the application process. Regular citizens can outsource the application process to be handled by the company which will eventually build the construction itself. (Evolta Oy, 2017a)

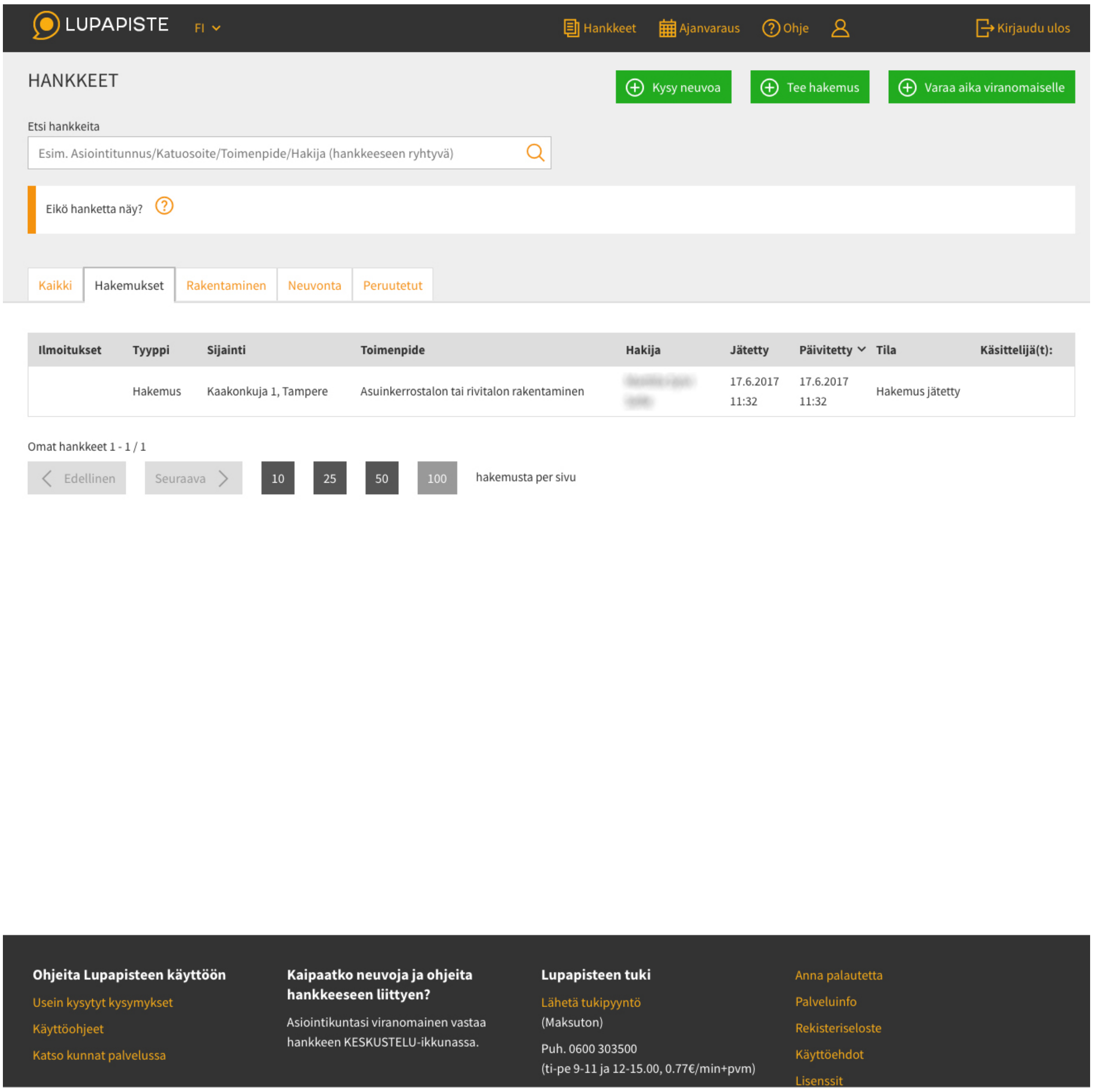


Figure 8: Lupapiste service - applications listing view

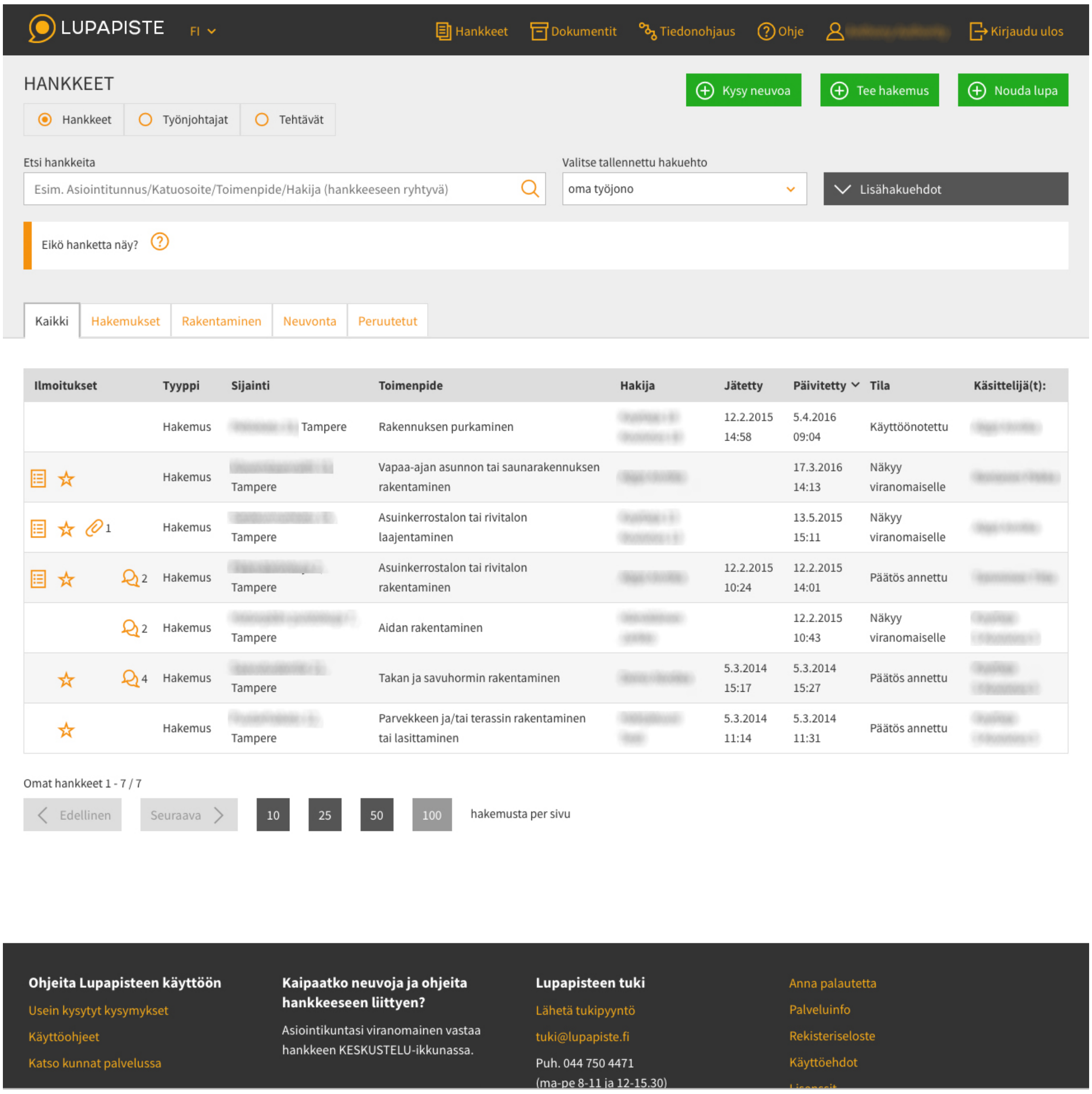


Figure 9: Application handlers, the official authority users view of the current applications

4.2. Research questions and methods

The aim of this thesis is to inspect the web analytics data and what can be learned from the users of the service. Multiple viewpoints were used when extracting and analyzing the web analytics data. Based on the analysis and research the objective is to be able to present actionable insights from the web analytics data. Found insights could be used to support the decisions made in the design process in the future and also to consider the upcoming developments to the service. Since the web analytics data isn't actively being utilized in the development of Lupapiste at the time of writing, this research also forms a baseline of the web analytics data. The baseline can be used in the future for comparing possible changes in the web analytics data for example when changes are made to a certain future. Through the case study presented later in the thesis I will research answers for the following questions:

RQ1: What can be learned from the users of the service in the e-government context with applied analytics?

- How the horizontal process of the service is visible through the analytics data?
- How to evaluate the current data collection implementation?

RQ2: How to evaluate the usefulness of the web analytics in the context of construction permit application?

- How the analytics data can be utilized in future improvements of the service?

As noted by Lazar et al. (2009), case study allows researchers to inspect one or more specific situations in a real life context by examining individual cases to build understanding of a certain behavior. Subjects of the case study should be carefully chosen in order to extract interesting and novel insights from the situations.

The case study in this thesis is done in an open-ended and explorative manner. Open-ended in the sense that the research question presented above are used as a starting point for the exploring of the web analytics data in Google Analytics. Explorative research in the context of this thesis means that if something interesting is found while exploring the data, these explorations and insights can be included in the results. Explorative nature is also present in each of the data gathering phases when, for example, different time ranges are used prior to data extraction and analysis.

The research was done through the Google Analytics user interface shown in Figure 10. For the most part of the data gathering and analysis, the standard reports in GA are utilized by exporting the data into a spreadsheet program for further transformation and analysis. The custom reports available in GA were also used for more sophisticated reporting. The process of the data gathering and analysis follows the web analytics process introduced in the section 2.2. By posing a question, the GA standard reports were first explored to found out whether sufficient data were found on the standard reports. When standard reports were found inefficient, the custom reports were created ad-hoc in order to explore in more detail. When sufficient data were found from the GA reports, it was then exported as a spreadsheet file into a spreadsheet software in order to analyze and visualize the data in order to find the answers beings research for.

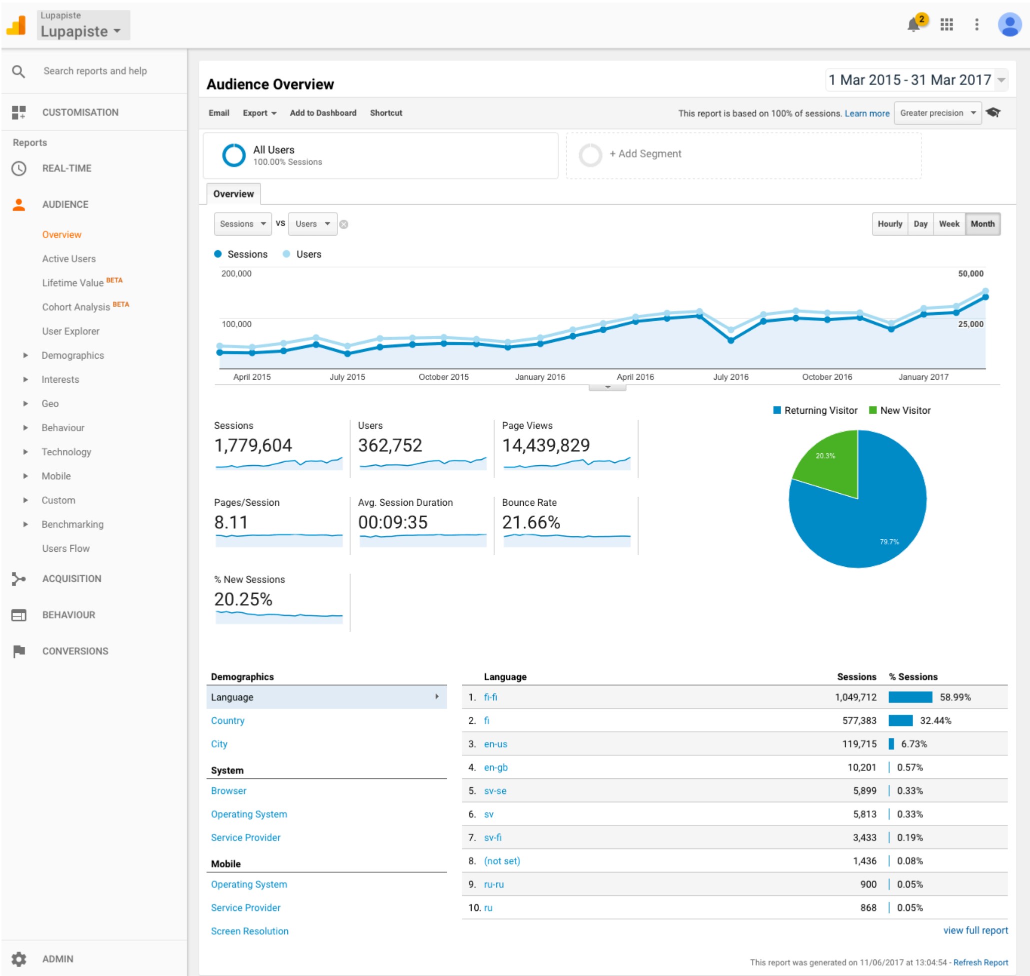


Figure 10: Google Analytics user interface

4.3. Instrumentation of web analytics

In this chapter I will examine the current Google Analytics implementation of the Lupapiste service.

As to why Google Analytics is used, there is no specific reason. Although, the developers and designers have earlier experience working with Google Analytics. The fact that Google Analytics is provided as free for all solution lowers the barrier of the implementation. Although, as noted by Nakatani and Chuang (2011), a free solution might not be always the viable choice. However, in the case of Lupapiste the free version of Google Analytics hasn't limited the utilization of web analytics.

In Lupapiste, the web analytics implementation utilizes the page tagging method described in section 3.4. However, the implementation is a bit more sophisticated since the tracking snippet is attached to various views with configuration files. Based on the inspection of the Lupapiste source code available in the GitHub[1], the tracking snippet is included on the main views accessed with different user roles. Inspecting the amount of

[1] https://github.com/lupapiste/lupapiste/blob/0fcab0488f17a2035826756ce7d0bdccc43a0542/src/lupapalvelu/components/ui_components.clj

pages captured by Google Analytics, between the 1st of March 2015 and 31st of March 2017, altogether 281 different pages are listed having page views. This amount contains only general pages in the service excluding pages with unique URL such as attachments.

In order to capture more fine-grained information about the user behavior, Lupapiste service has been implemented with the Google Analytics event tracking. Event tracking in *GA* enables detailed tracking of the user interactions with the content on the web site. Events in *GA* contain following parameters: event category, event action, event label (optional) and event value (optional). Events are categorized by using the event category which can be used in multiple events to group relating interactions together. Event action is used for naming the event or user interaction. For example, “Videos” event category’s event actions could be named as “Play”, “Stop” and “Pause”, indication distinctively the user behavior. The event label adds another level of detail to describing the interactions. Event labels can be used, for example, to give a title to a video watched by the user or to name a file user downloads from the website. The event value can be used for mediating an integer value associated with the event which can be used, for example, to give a monetary value to an event. (Google, 2016)

In Lupapiste, the current event tracking implementation captures events in eight distinctive categories: Application, Applications, Attachments, Conversation, Create, Inforequest, Side-Panel and Tree. Each of the category relate to a certain context within Lupapiste service. The amount of event actions capture currently contains 72 event actions, while there are 159 distinctive event labels found currently through Google Analytics. The events are discussed in detail later in the chapter 5.5.

Based on the amounts of categories, actions and labels, it is not straightforward to evaluate the coverage of the current implementation. To assess the details of the coverage would require extensive research of the source code of Lupapiste in order to see whether there are some aspects of the user interactions missing from the implementation. Furthermore, the coverage of the web analytics could be evaluated more efficiently based on an event tracking documentation in which all the views, functions and attached events would be described. The current event tracking has not been documented extensively making it difficult to see if there are user interactions not being tracked with events. In the future, the event tracking documentation would allow extending the current event tracking in cases where new views and functions are added to Lupapiste. This approach calls for the web analytics to be considered already in the designing phase in the sense that the event tracking is planned and documented extensively in the future.

5. Results

Following section concentrates on the results and analysis of the case study research performed on the Lupapiste web analytics data. Extraction and reporting of the web analytics data is done through different viewpoints of the data. Various reports and extractions of the data were performed to analyze the user behavior of the Lupapiste service.

5.1. The number of users and sessions across a timeline

Without prior knowledge about the number of users concerning the Lupapiste service, this section explores what temporal frame should be applied as a filter to the web analytics data.

The inspection was started by finding what is the starting point of data capture by manually browsing the Overview section in Google Analytics. Based on the exploration, 1st March 2015 is the first day from which data has been captured. Since the whole lifetime of the analytics data is the scope of the inspection, the last possible month is used as the end of the time period ending at the 31st March 2017. With this timeframe, it is possible to explore data from 24 months altogether.

Month	Sessions	Users
3/2015	32363	11242
4/2015	31596	10650
5/2015	35351	12589
6/2015	47626	15354
7/2015	29719	11226
8/2015	42806	14949
9/2015	47826	15184
10/2015	50035	15466
11/2015	49343	14548
12/2015	42631	13110
1/2016	49352	15308
2/2016	64188	19214
3/2016	77001	22290
4/2016	93164	25506
5/2016	99233	27364
6/2016	103998	28200
7/2016	55774	19143
8/2016	93439	26698
9/2016	99602	28474
10/2016	96640	27501
11/2016	100687	27570
12/2016	78075	22383

1/2017	107291	29719
2/2017	110569	30770
3/2017	141295	38247

Table 3: Monthly amounts of users and sessions

Based on the inspection of the users and session amounts, both of the metrics have been increasing throughout the timeframe. In order to inspect the details of the users and session amounts, the data were exported to a spreadsheet software and transformed into graphs discussed in the follow section.

Shown in Figure 11, the amount of users per month within the time period is presented. The number of users has been steadily rising throughout the time period, peaking at the March 2017. What is interesting are the dips in the data: December 2015, July 2016 and December 2016. These months contains a lot of holidays which would explain why there is a visible decrease in the number of users. The trend in the number of users was explored by adding a trendline to the graph shown in Figure 11 below. What is notable is the raising trend in the number of users within the timeframe.

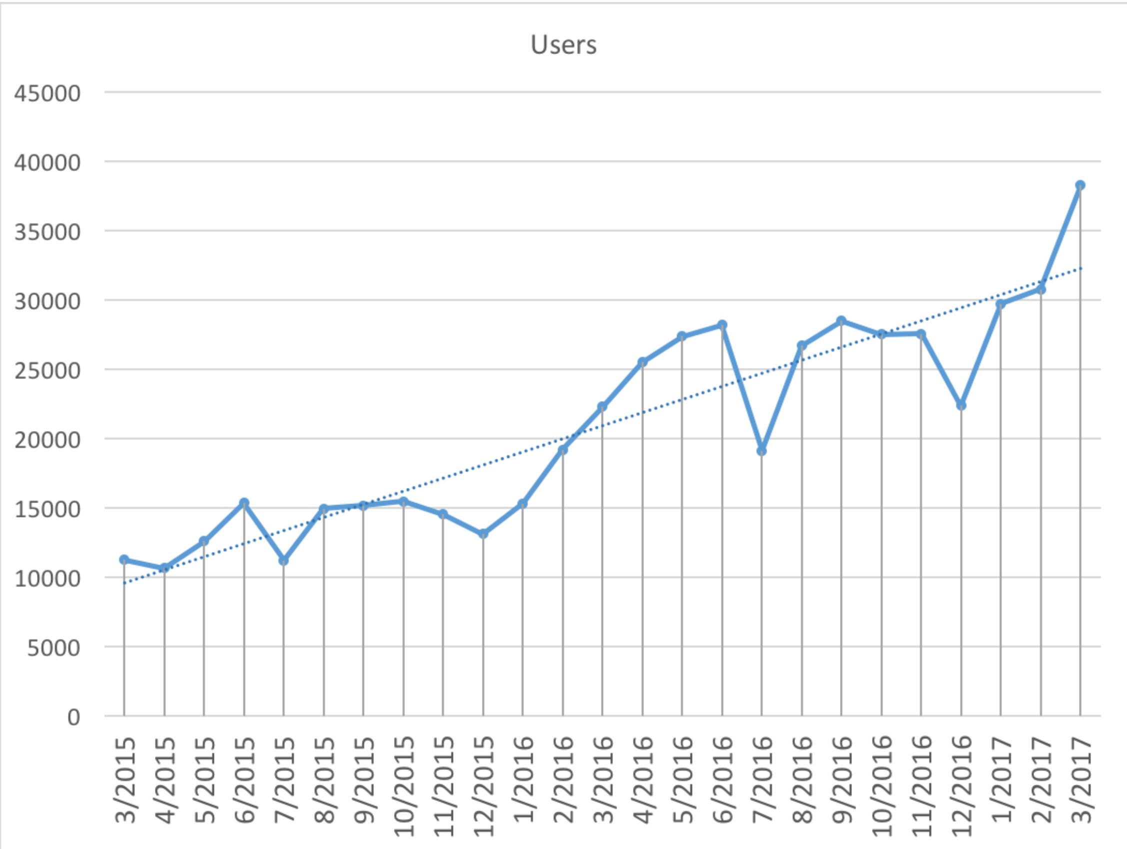


Figure 11: Number of users between 3/2015 and 3/2017.

In the following Figure 12, the number of session per month are shown from the time period. In Google Analytics, sessions cover all the interactions within one session which starts from the initial hit and is active until thirty minutes of inactivity. If there are interactions after the 30 minutes of inactivity, this counts for 2nd session. (Google, 2017b)

Plotting the data of the number of sessions, the graph resemble the number of users shown in Figure 11. Number of session also include the same decreases with the amount of session as with the number of users shown in Figure 12.

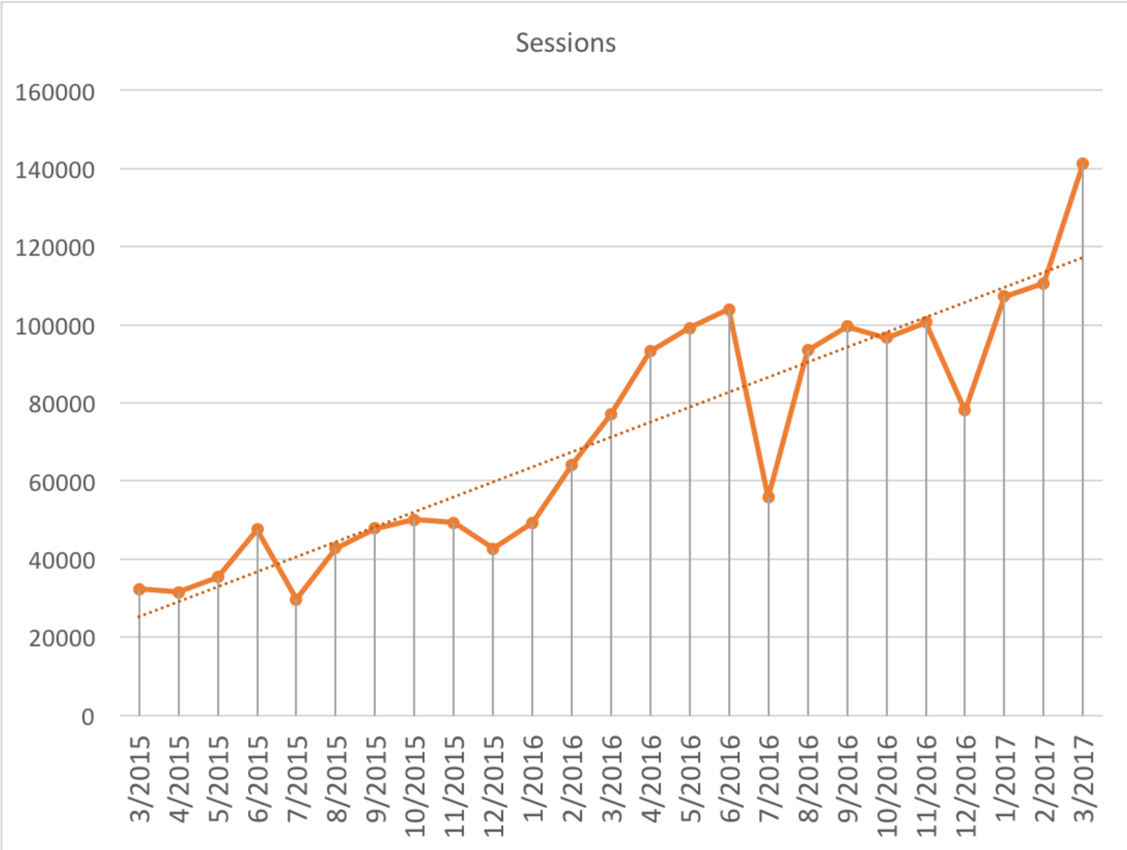


Figure 12: Sessions between 3/2015 and 3/2017

What is visible in both of the graphs (Figure 11 and Figure 12) is the raising trendline in both of the graphs. This is a natural progression explained by the fact that more and more cities and municipalities have started to use the Lupapiste service. From the UX designer's perspective and also from the business perspective, the trend of raising in both metrics is more insightful, meaning that the utilization of the service has been successful in cities and municipalities. While the raising trend might be a natural assumption, the raising trend can be verified through the web analytics data.

As the number of users and sessions are default metrics available in GA, the number of users might be more natural number to report as it relates to the notion of users, essentially meaning real people using the service. On the other hand, the number of sessions enables the inspection of the usage in a detailed level.

5.2. Returning versus new users

Since the number of users have increased during the time frame (3/2015-3/2017), this begs to question what is the proportion of new users versus users returning to Lupapiste. Across the time frame from 1st of March 2015 to the 31st of March 2017, 79.70 percent of the users are returning users and 20.30 percent are new users shown in Figure 13 below.

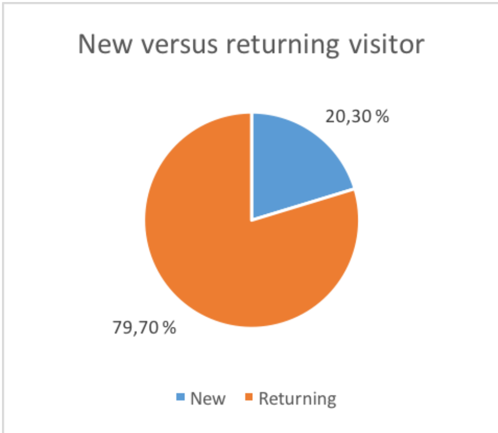


Figure 13: New visitors versus returning visitors

The majority of the returning users might be explained by the fact that the permit process requires the users to come back to the Lupapiste service. The process requires multiple visits to Lupapiste as the permit requires different information in different steps of the process.

5.3. Usefulness of segmenting the web analytics data

As noted by Beasley (2013) segmenting web analytics data based on the user behavior is where web analytics excels as a tool. Segmenting data enables the exploration of data in order to find patterns and finding more specific answers to posed questions. (Beasley, 2013)

With Lupapiste, users can be divided into two distinctive segments based on the user types: applicant users and authority users. Applicant users represent the users applying for the construction permit while the authority users are the officials handling the applications in Lupapiste. The two user groups have different features and function available in the service thus it is natural to inspect data through the segments. Furthermore, the segmentation of the data can be based on the differences in the URLs of the service. When the applicant users log in to the service, the URL will include "/applicant/" section. The URLs of the authority users will include "/authority/" section.

In order to inspect the data based on the segments, segments were added with the use of the Google Analytics administrative interface. Both of the segments utilizes the Page condition matching a regular expression: */app/[a-z]+/applicant/.** for the applicant segment and */app/[a-z]+/authority/.** for the authority segment.

The question "What can be found out from the analytic data if when segmentation is applied to the data?" was used as a starting point of utilizing the segments in extracting the data from GA. Shown in Figure 14, are the number of users of Authority and Applicant segments with the overall number of users. As shown in the below Figure 14, the number of users in the Authority segment is much lower compared to the number of users in the Applicant segment. Overall the same holiday months are also present, especially in the Applicant segment. However, in the Authority segment the dips are not as drastic.

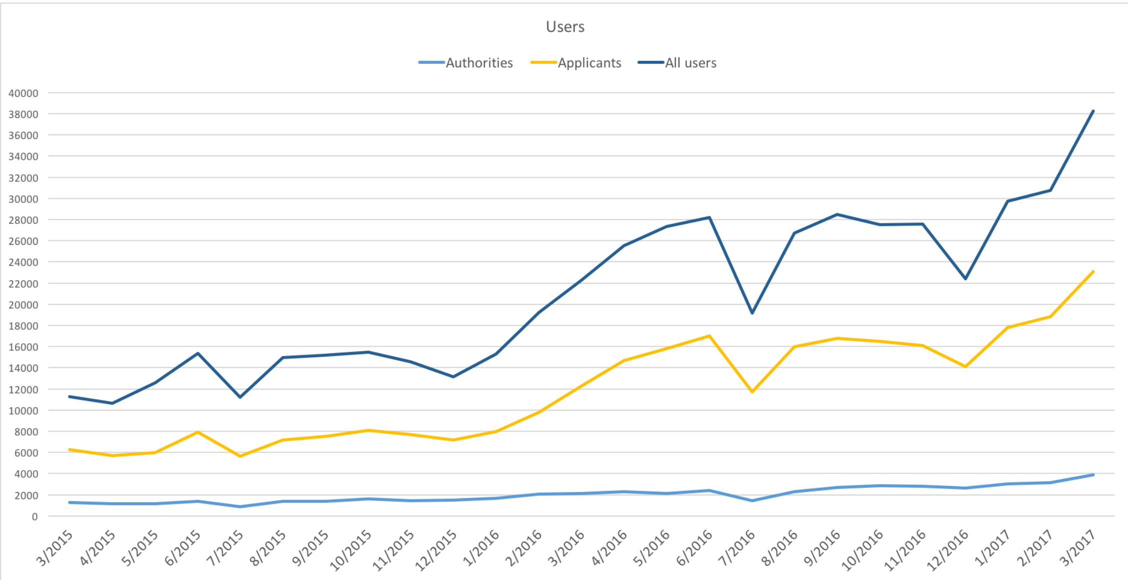


Figure 14: Number of all users and users per segment

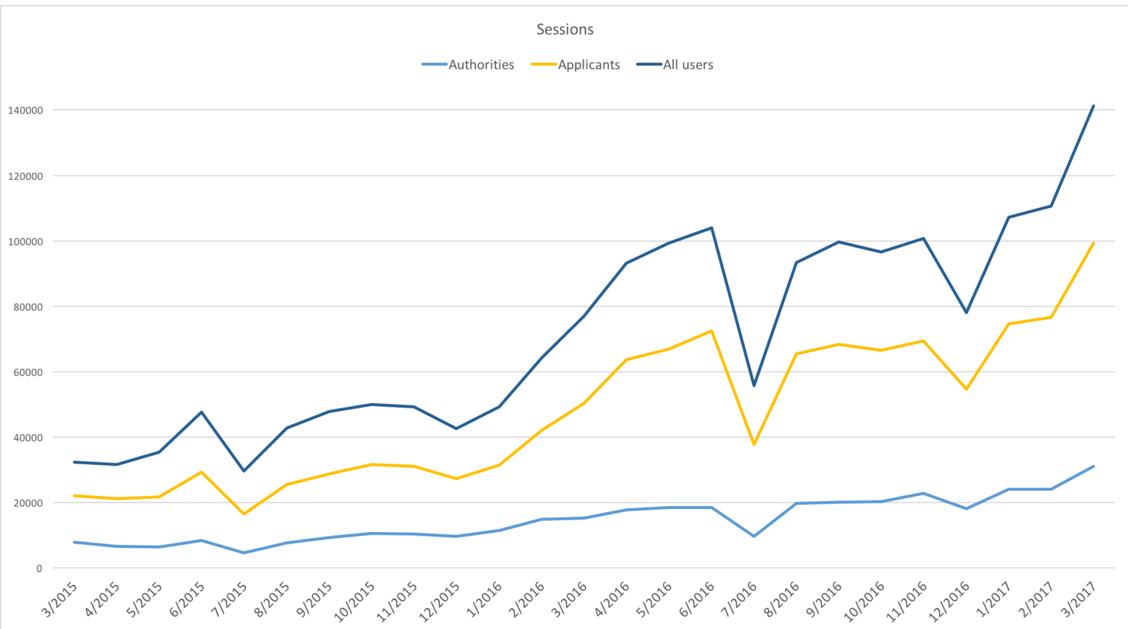


Figure 15: All sessions and segmented sessions

What is visible in both Figure 14 and Figure 15 is that Applicants segment includes more users and therefore more sessions compared to Authority segment. The phenomena behind the differences in the segments is explained by the fact that there is simply more people applying for construction permits than authorities handling the applications. More importantly it is interesting that there is a steady increase in both of the segments in the amount of sessions and users. The raise in the metrics reveal the expanding coverage of the Lupapiste service thus both of the segments are showing increase.

From the development perspective, a sudden surge in either of the segments number of sessions or users might be an alarming sign which should be further investigated in order to find an explanation behind the surge, for example technical issues. The trends in the number of users and session might be more useful to inspect and report regularly to indicate whether the maximum number of users have been reached based on the coverage of the service.

5.4. Usage hours during the days of week

In this section, the usage times of the Lupapiste service during the week and hour of the day are inspected in detail. The segments of the Authority users and Applicant users are utilized in order to see if there is distinctive behavior with in the two segments.

As Lupapiste service is accessible 24/7, especially from the applicants’ viewpoint this enables using the service, thus applying for the permit, whenever it is suitable for the users. To research if this is present in the data of the Applicant segment, the usage times of the service based on the day of the week and hour of the day were extracted and transformed into a graph shown in Figure 16.

What is interesting is the fact that Applicant users are using the service mostly on Monday through Friday, as visible in Figure 16. During weekends only a small portion of the users are accessing the service. Interestingly Sunday is more active day compared to Saturday. However, beginning of the week through Tuesday, Monday and Wednesday, are the most active days, followed by Thursday and Friday.

Examining the Figure 16 further, Applicant users are most actively using the service starting 7 o’clock in the morning until 17:00, which quite distinctively follow regular office hours. Though there are some users accessing the service later in the evening between 18:00 until 21:00. The beginning of the day from 01:00 until 06:00 are the quietest hours during the day. It is also interesting that there is a big drop-off between 11:00 and 12:00 in the usage of the service, suggesting that Applicant users are having a lunch break. Since the most of the usage happens during the office hours, Applicant users are perhaps using Lupapiste while at work.

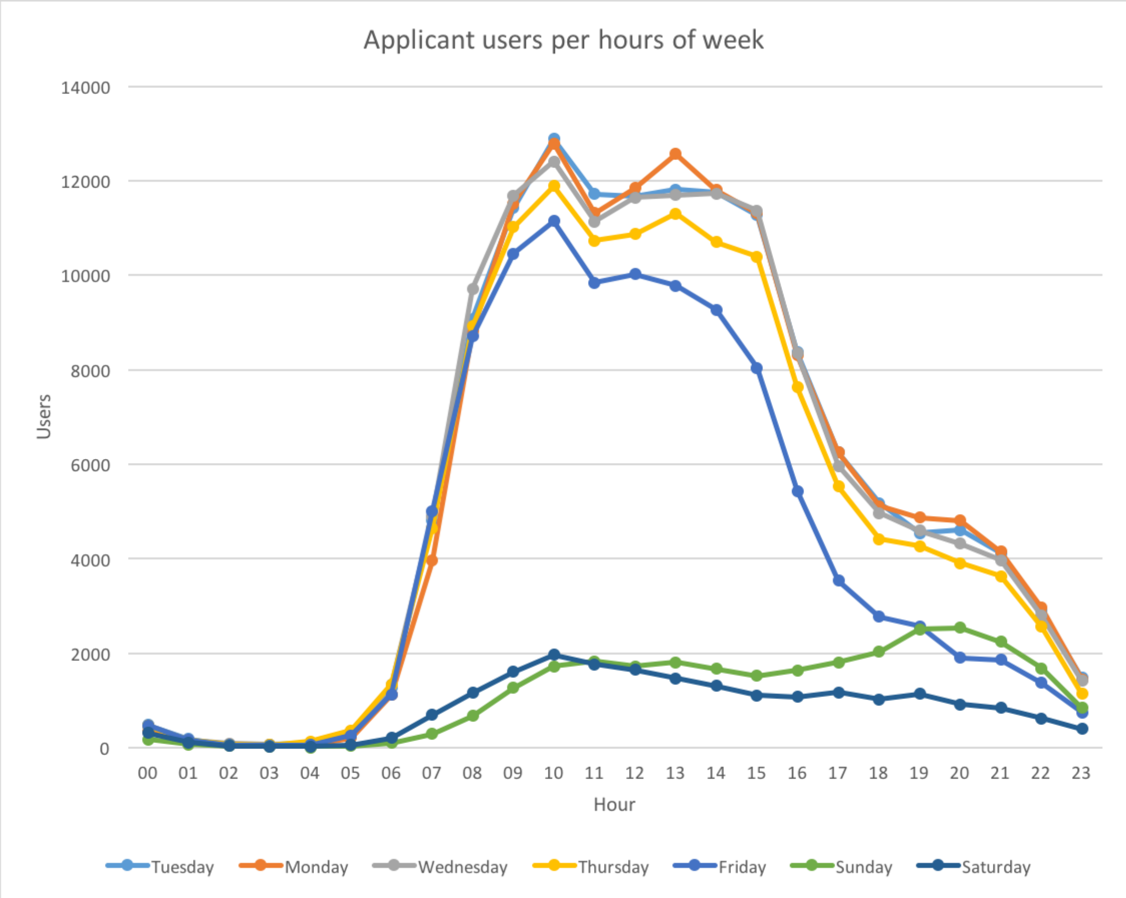


Figure 16: Applicant users in week days per hour

I next set out to find, what kind of usage patterns are found within the Authority user segment. Is there clear differences in the usage days or hours during the days or does the patterns follow the Applicant user segments patterns. In order to find out answers, data was extracted from the Google Analytics into spreadsheet program and transformed into a graph presented in Figure 17.

What is distinguishable in Authority user segment compared to the Applicant user segment is Authority segment's usage is more restricted to the office hours starting from 07:00 through 16:00. The usage after office hours is not present at the same volume as with the Applicant users segment. Most of the Authority segment's users are active before lunchtime from 07:00 until 10:00 and after the lunchtime from 12:00 until 16:00. The lunch hour, at 11:00, shows visible drop-off with the Authority users visible in Figure 17.

Examining the usage based on weekdays, the Authority user segments activity during the weekend on Saturday and Sunday is essentially non-existent as opposed to the Applicant users in Figure 17. Both Saturday and Sunday following the same curve in the graph (Figure 17). Interestingly there is more usage during Sunday, perhaps suggesting Authority users might be preparing or examining the applications in advance for the upcoming week.

Inspecting the usage based on the day of the week, Tuesday is the most active day in the Authority user segment, Thursday and Wednesday following second and third. Interestingly Monday is the fourth active day. What is also interesting is the usage during Friday is much lower compared to Monday or other days of the week shown in Figure 17.

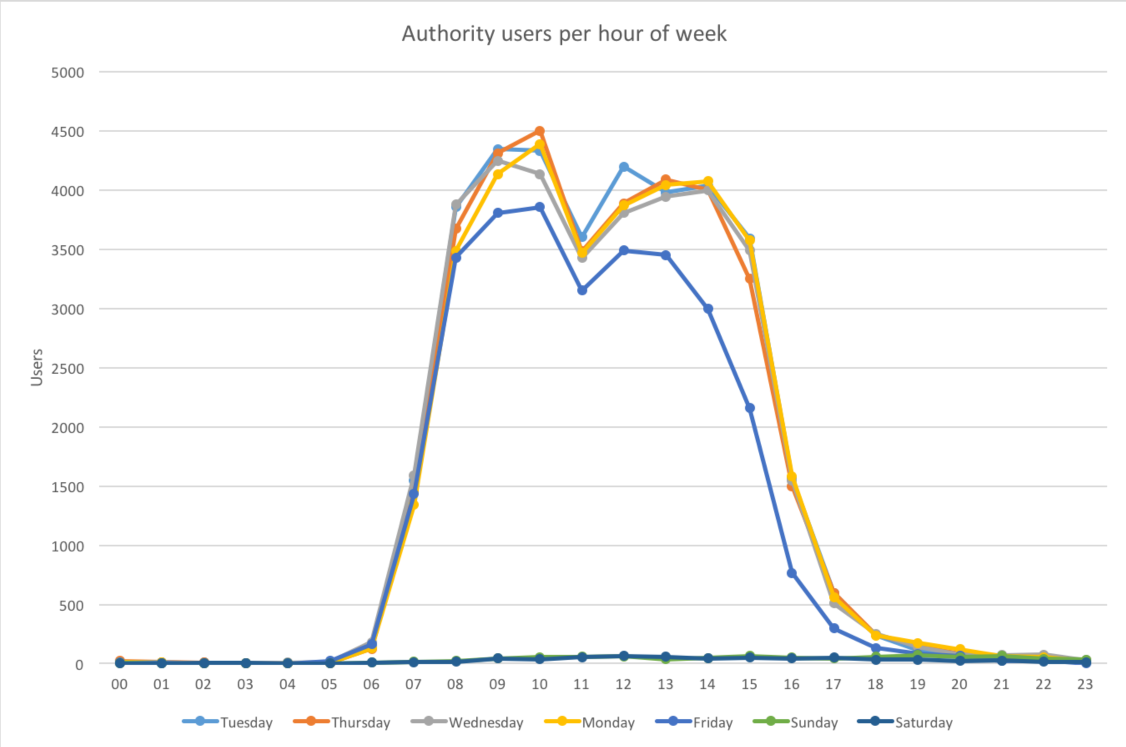


Figure 17: Authority users during week days per hour

Even though the Lupapiste service is available regards of the office hours, still the usage follows the common workday starting around 07:00 and ending around 16:00 in both of the segments. On the other hand, the Lupapiste customer service is available during the office hours, which might explain the phenomenon. The Authority segments usage times indicates the usage of the Lupapiste service is embedded into the day-to-day work process. On the other hand, the Applicant segment's users are more inclined to using the service after office hours late in the evening.

5.5. Usage patterns based on event tracking

The current implementation of web analytics in Lupapiste captures events in eight different categories listed Table 4 with additional description of the context of the categories relating to the service. As discusses earlier in the chapter 4.3, events in Google Analytics are broken down into event categories, event actions and event labels.

Event category	Context of events

Application	Events which are fired when a single application is viewed.
Applications	Events which concern the applications listing view.
Attachments	Application's attachments which are manipulated in the context of a single application.
Conversation	Events captured when users engage with the conversations relating to an application.
Create	Events relating to the creation of a new application.
Inforequest	Events relating to the information request context which can be created before creating the actual application.
Sidepanel	Events initiated when users open the help view in the application context.
Tree	Events relating to the navigational tree element which is used in creating new application or information request.

Table 4: Event categories and contexts

The Figure 18 below, presents a more detailed view of the event tracking used currently in Lupapiste by examining the number of event actions in each of the event categories. The Application context has most event actions (34) attached to the different functionalities. Attachments context is equipped with 12 event actions, followed by the Create context and Applications context. Based on the Figure 18 and the contexts of the event actions, the event tracking seems quite extensive. Contexts which contain a lot of functionalities have also many event actions attached.

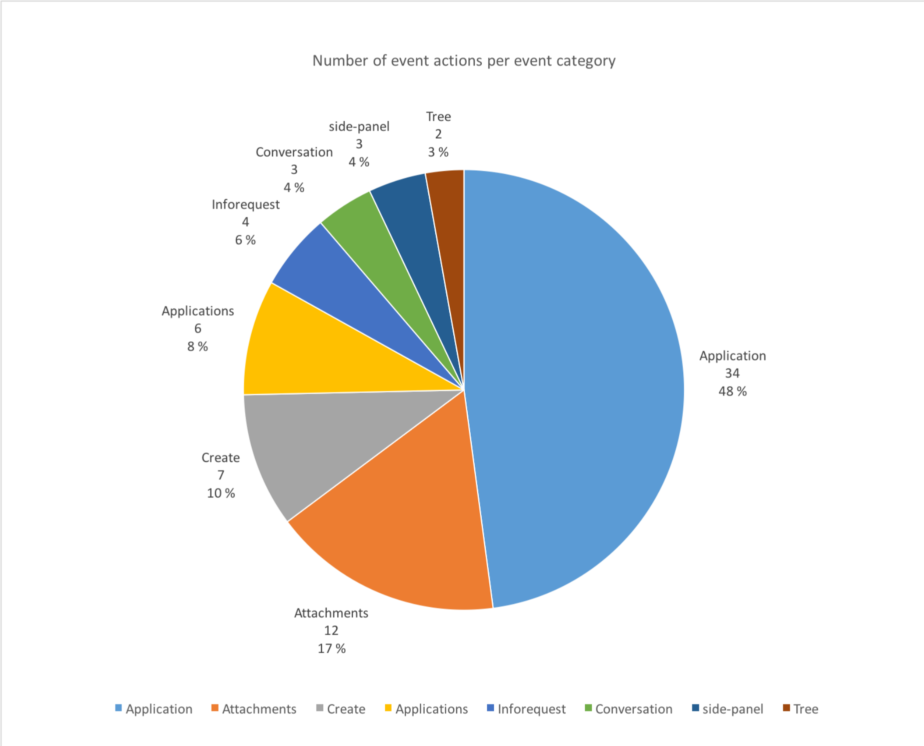


Figure 18: Number of event actions in the event categories

Inspecting the events from the action level, disregarding the event category, are shown in Figure 19. Based on the Figure 19, the current event tracking in Lupapiste currently captures 71 different event actions altogether. The graph below is based on 10,334,458 total events and 4,734,683 unique events. Overall, the event actions relating to the navigation in the Lupapiste service are highlighted in the graph (Figure 19) representing actions which are triggered the most by the users. The first six events containing the most events, shown in the Figure 19, namely *changeTab, openApplication, radioTab, backToApplication, nextTab* and *treeBack*, are all events which are triggered when users move between different views in the service. *changeTab* event action is triggered when navigation tabs are changed within the opened application. The *openApplication* event action is triggered every time an application is viewed. The *radioTab* event action is triggered when user is browsing the different types of applications in the application listing view which if the default view shown to the user when logging in. The *backToApplication* event is triggered when a user returns to the application view from the application's attachments view. The *nextTab* event relates to the creation of the application, triggered when a user navigates to the next part of the application. The *treeBack* event is triggered when users go one step backward in the application's subject selection tree.

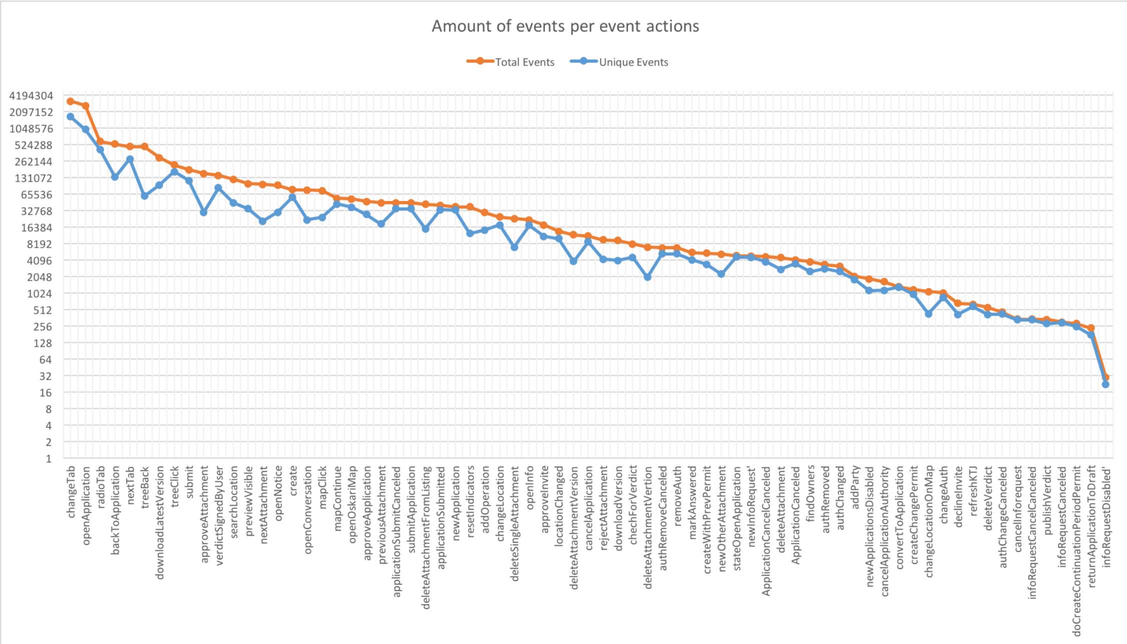


Figure 19: Events across different event actions

As mentioned earlier, evaluating the thoroughness of the event tracking is not straightforward process and would require more time, effort and technical skills in order to fully understand how complete the current event tracking implementation actually is. Therefore, the analysis of the event tracking is shallow and concentrates on the amount of events.

In order to give some context to the even actions, the event flow of creating a new construction application is following:

Category: Applications, Action: create, Label: create

Category: Create: Action mapClick, Label: map

Category: Create: Action mapClick, Label: continue

Category: Tree, Action: treeClick, Label: Rakentaminen ja purkaminen

Category: Tree, Action: treeClick, Label: Uuden rakennuksen rakentaminen

Category: Tree, Action: treeClick, Label: kerrostalo-rivitalo

Category: Create, Action: newApplication, Label: tree

Category: Applications, Action: openApplication, Label: application

Furthermore, when users create a new information request in the Lupapiste service, the flow of the event categories, event actions and event labels is the following:

Category: Application, Action: create, Label: create

Category: Create, Action: mapClick, Label: map,

Cateogyr: Create, Action: mapContinue, Label: map,

Category: Tree, Action: treeClick, Label: Rakentaminen ja purkaminen

Category: Create, Action: newApplication, Label: tree

Category: Create, Action: newInfoRequest', Label: tree

Category: Applications, Action: openApplication, Label: inforequest

Furthermore, when applying the Authority and Applicant segments to the event data, it is possible to see how the users in these two different segments are using the service and what kind of differences there might be between the segments.

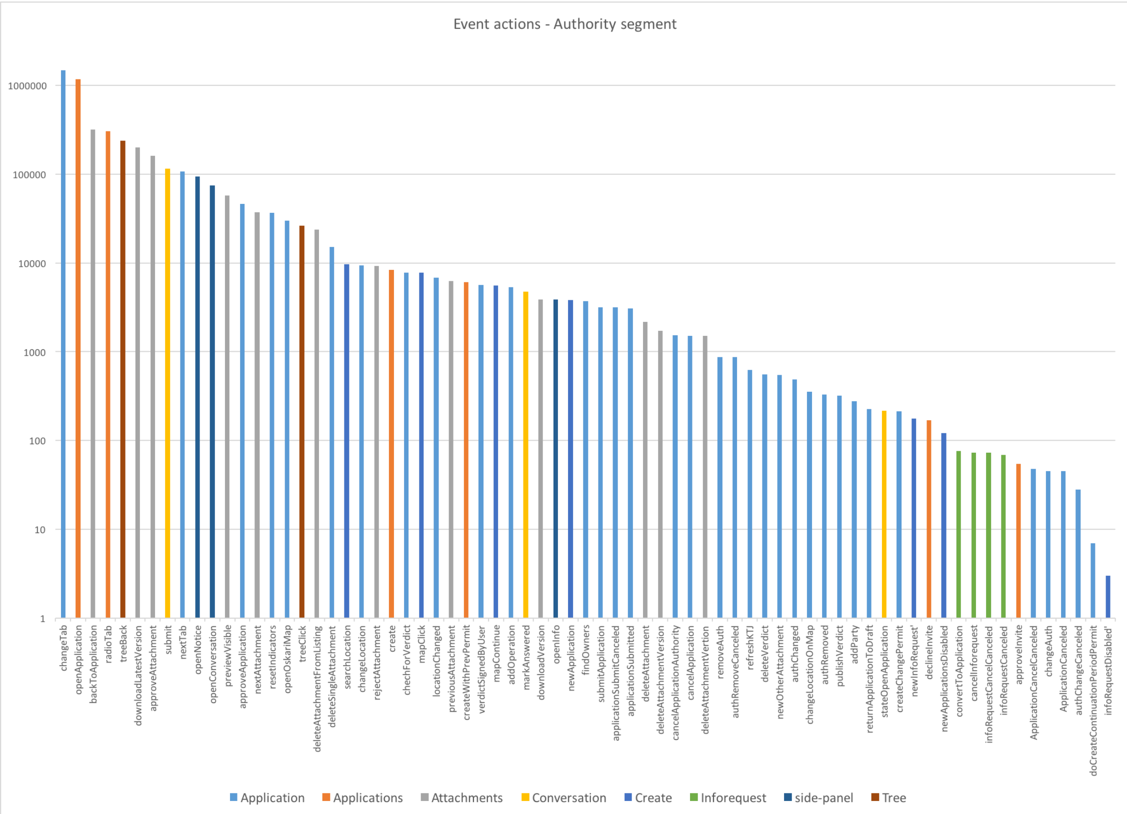


Figure 20: Authority segment's event actions

The above Figure 20 presents Authority segment’s the number of event actions in the different event categories. The high value of *TreeBack* event action raises questions about the efficiency of the navigational tree component. This brings up the question of are users having hard time to decide when navigating between the different states in the subject tree because they end up pogo-sticking between the different states? The issue of the navigation, especially the subject tree element, should be inspected in further in usability studies to see whether the navigation tree is hindering the user experience.

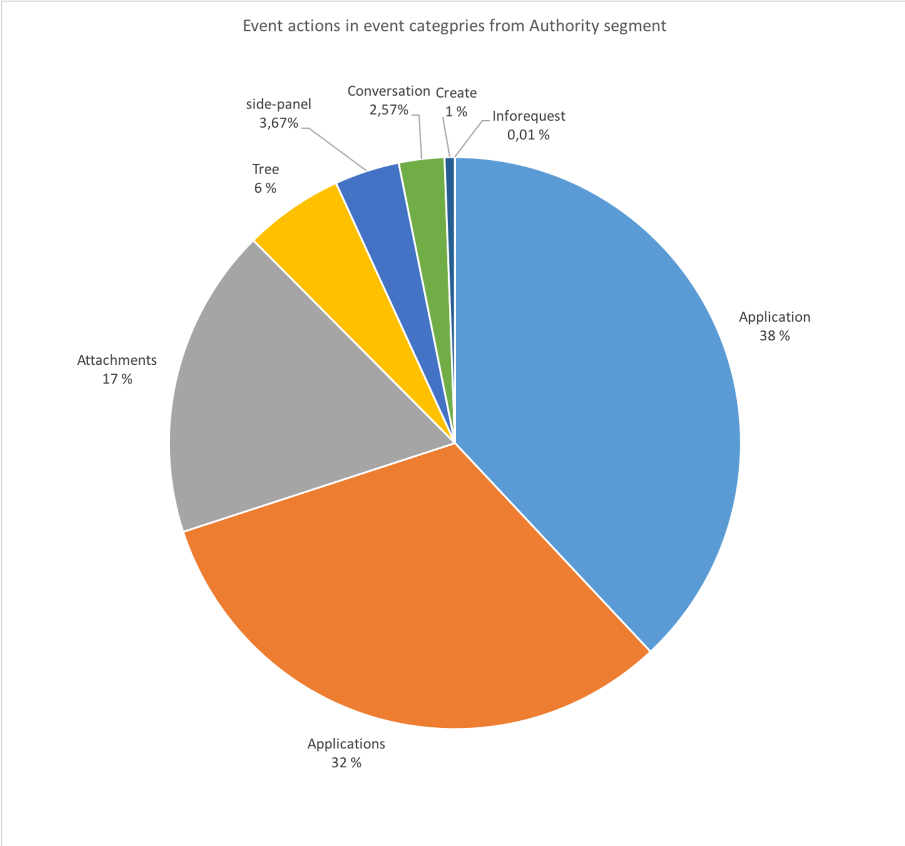


Figure 21: Event actions in event categories from Authority segment

The above Figure 21 presents Authority segment's event actions in the context of the event categories. Based on the pie chart shown in Figure 21, most of the event actions of the Authority segment are triggered in the Application context (38%), followed by the Applications context (32%) and Attachments context with 17 percent share. Interestingly event actions triggered in the Tree context has 6% share of the event actions.

When considering future development efforts, the above Figure 21 could be considered as a guideline in targeting the development, especially for the Authority segment. As most of the events take place in the Application context, the highest value of the improvements for the Authority segment users could be gained by developing the aforementioned section of the Lupapiste service. Also, when considering what sections of the service should be tested in usability studies, the Figure 21 could be also used for planning and designing the usability studies.

The Figure 22 below presents the event actions of the Applicant segment. The navigational event actions are similarly visible with the Applicant segment compared to the Authority segment. As shown in the Figure 23, most of the events (44%) happen in the Application context, followed by the Applications context (33%). Interestingly the Attachments context has 9% of event actions, while the same context with Authority segment has 17% of the event actions.

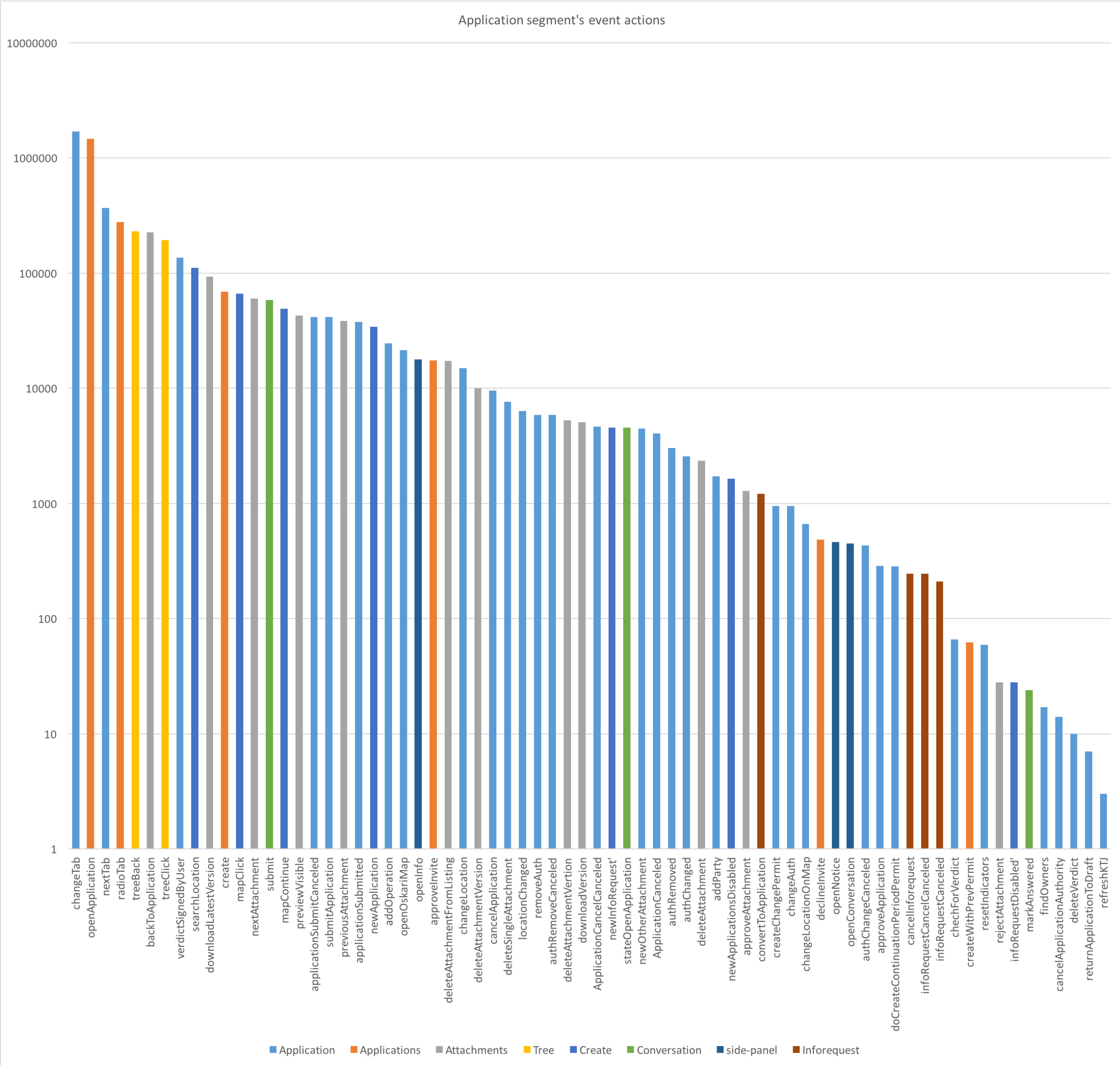


Figure 22: Applicant segment's event actions

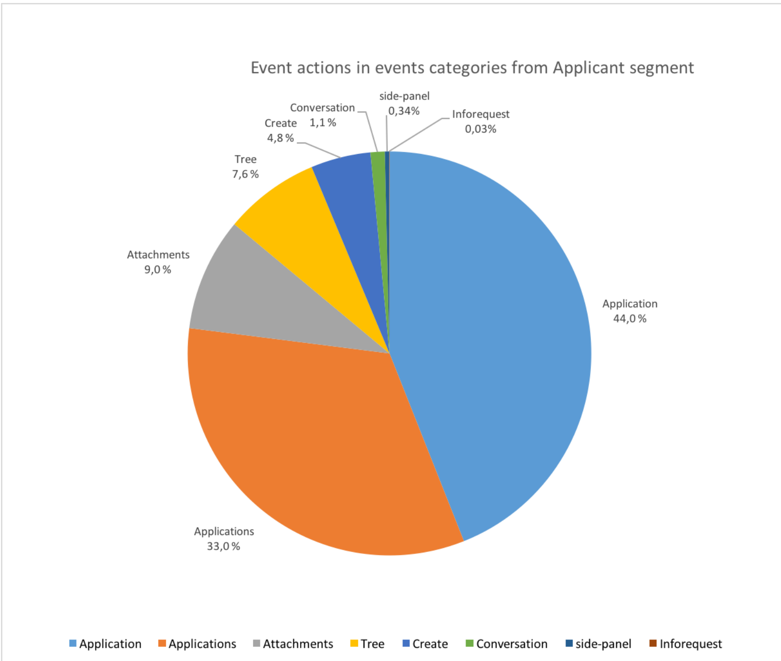


Figure 23: Applicant segment's event actions in event categories

Segmented event data should be evaluated when considering further improvements to the service. The events, depending on the detail of instrumentation, give a concrete picture of where the interactions take place in the Lupapiste and which interactions are used the most. The segmented event tracking data enables inspection of the possible differences and similarities between the two user segments. In the case of Lupapiste, the context of the event actions in both of the segments are quite similar. Based on this information, when future improvements are made to the service, users in both of the segments will benefit from the enhancements.

5.6. Event flow from analytics

The Behavioral Flow in the Google Analytics enables one to inspect in which order the events occur.

The below Figure 24 and Figure 25, presents the flow of the events of the Authority segment (Figure 24) and the Applicant segment (Figure 25). Event flow is restricted to the first four events occurred in the flow. Based on the figures, the flow of events does not contain clear and linear paths but highlights the pogo-sticking behavior of the users. A user opens an application, views a certain section within the opened application and then returns back to the applications listing view.

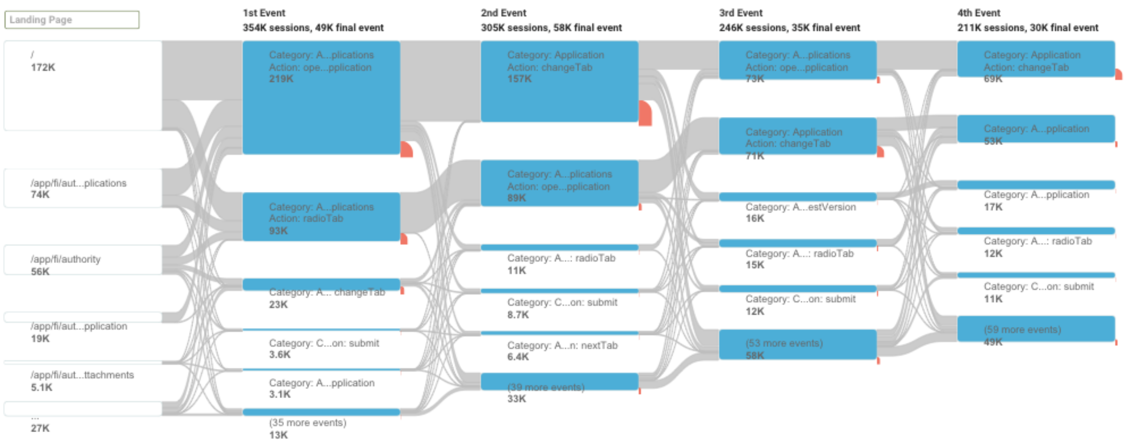


Figure 24: Behavioral flow of the Authority segment

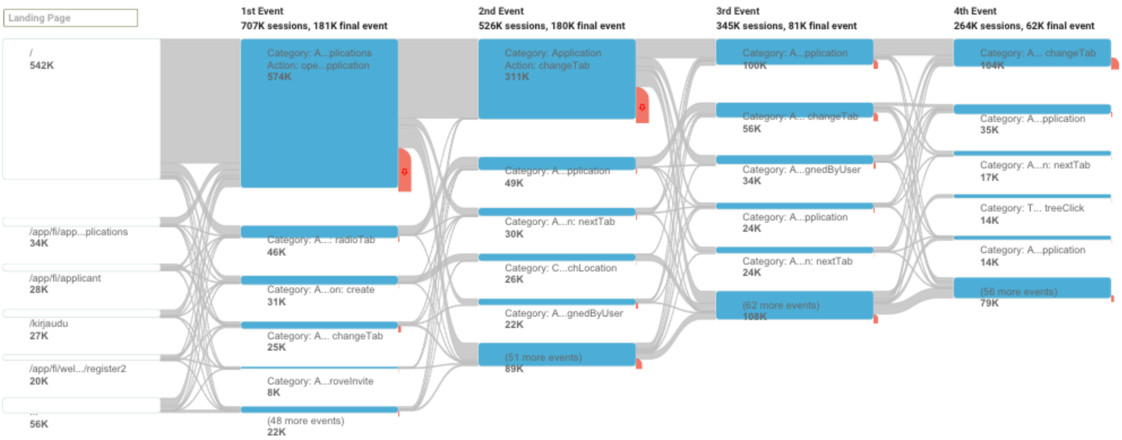


Figure 25: Behavioral flow of the Applicant segment

5.7. Device usage

As more and more users are utilizing different devices to access websites, this warrants for design decisions which take it into account that users might not be only using a desktop computer with a big screen and mouse and keyboard as the main input methods. To evaluate whether the Lupapiste service should be optimized for a specific device category, following section inspects the web analytics data based on different device categories.

Table 5 contains the amounts of sessions based on categorization of devices used for accessing Lupapiste. The desktop category represents the biggest portion with 86 percent share of the device categories. Interestingly mobile category is bigger device category (9%) than the tablet category, which includes only 4 percent sessions based on the Figure 26. The short average session duration in mobile and tablet categories might be explained by the fact that users access the service just to take a peek whether there have been updates or if there is something new that the users should react upon. Comparatively high bounce rates in the mobile and tablet categories suggest also this behavior. Still it is interesting that more new sessions come from mobile and tablet devices compared to the desktop device category. Users might use tablet or mobile devices as an additional way for accessing the website in addition to the desktop devices.

Device Category	Sessions	Pages/Session	Avg. Session Duration	% New Sessions	Bounce Rate
desktop	1647129	8,74	10:35	18,30 %	19,71 %
mobile	178464	3,76	03:14	31,03 %	34,31 %
tablet	88360	5,40	05:25	32,97 %	31,36 %

Table 5: Device usage across desktop, mobile and tablet device categories

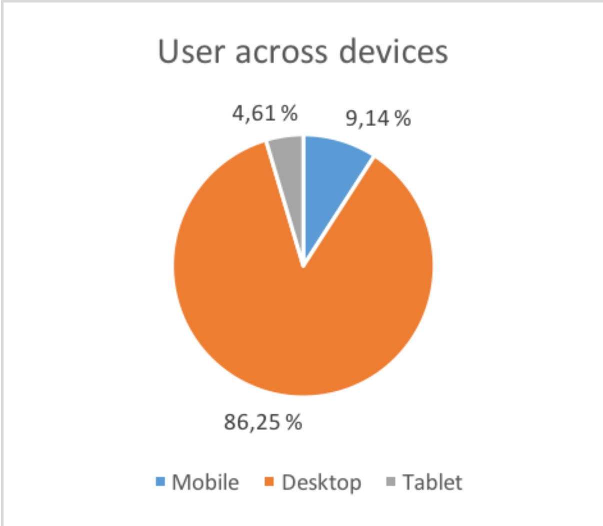


Figure 26: Users across devices categories

Based on the previous Table 5 and the above Figure 26 about the differences between the different devices used for accessing the service, a clear view of the device spectrum of Lupapiste service is formed. But more interestingly, what kind of changes there has been if we inspect the different devices categories across the timeline? In order to reveal possible differences in the device categories, sessions data from the device categories was extracted and plotted in spreadsheet software.

Based on the Figure 27, we can see that the Desktop category has been increasing throughout the timeline, which is expected as the amount of users overall has been increasing with a same rate based previous Figure 14 and Figure 15. Based on the linear trendline, the share of the desktop device category will increase in the future as well.

Inspecting the other two categories in more detail it is visible that far fewer sessions are being initiated from the mobile and tablet devices. The mobile device category covers more sessions than the tablet category. Based on the device market shares in Finland, shown in Figure 28 (StatCounter, 2017), the mobile device category is expectedly bigger than the tablet device category.

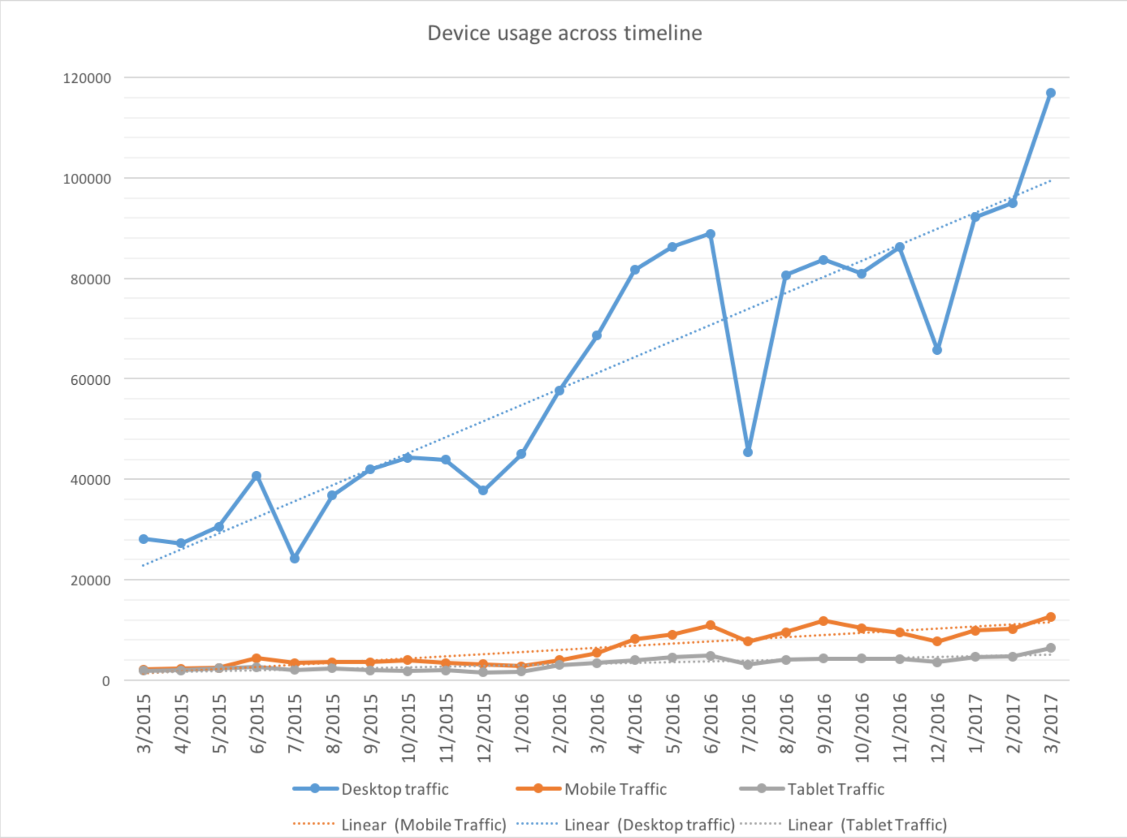


Figure 27: Device usage development across timeline

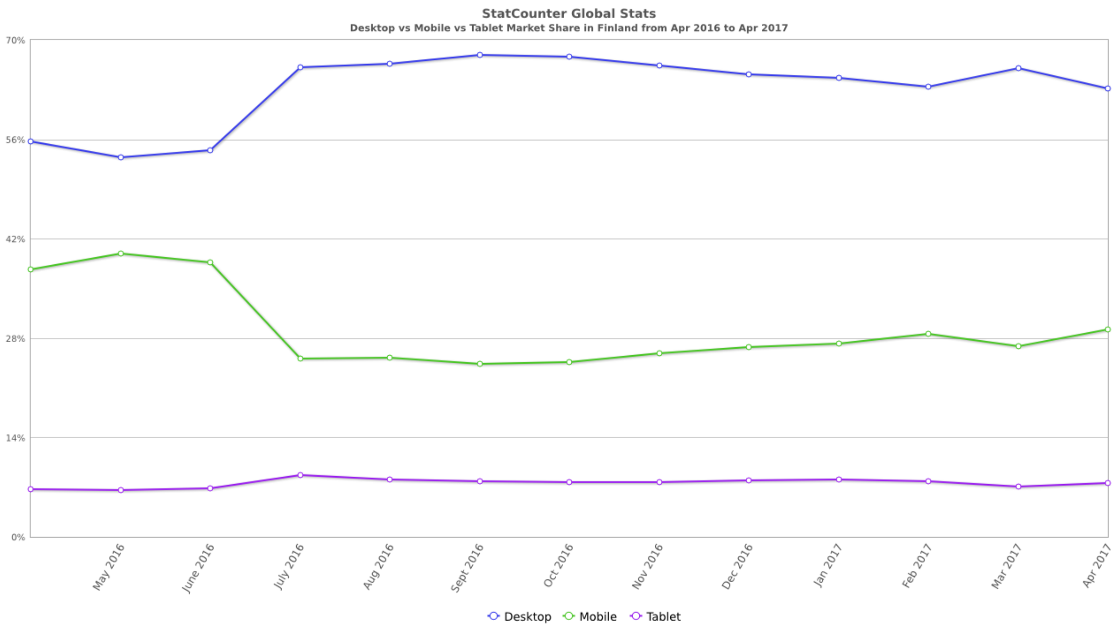


Figure 28: Desktop vs Mobile vs Tablet Market Share in Finland (StatCounter, 2017)

Based on the previous section, from the user experience designer's point of view, it is important to acknowledge the differences between the device categories when considering future improvements and developments to the Lupapiste service. Optimizing the service for smaller screen might not be efficient business decision as the session amounts are quite low compared to the desktop devices. Also, the sessions from the mobile and tablet device categories are increasing but in mild pace, which might change in the future. The evolution of the device categories should be inspected from time to time in order to see whether there have been such changes in the growth of the device categories to reconsider the optimization for other devices as well.

5.8. How users arrive to the service

As Lupapiste service is available in http://www.lupapiste.fi URL, it might be tempting to rely on the fact that users remember this quite simple address and type it in to their browser. From web analytics, it is possible to see what channels users are using when arriving to the service. In analytics terms, channels refer to the way user have reached the service. For example, an organic channel means a user utilized search engines results to discover and access the service.

Based on the following Figure 29 and Figure 30 users in both of the segments arrive through the direct channel in most cases, meaning users either type in the URL or utilize bookmarks from their browser. Referral is the second popular channel in both of the segments, closely followed by organic search channel. Essentially users are directed to the service from another service by a link. What is noteworthy is the portion of users using search engines in arriving to the service.

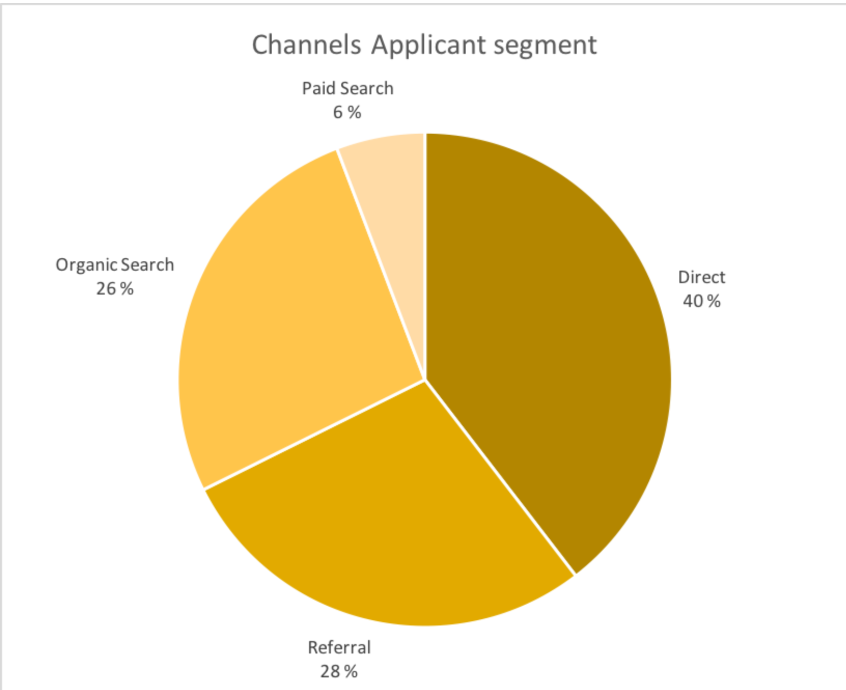


Figure 29: Channels based on the Applicant user segment

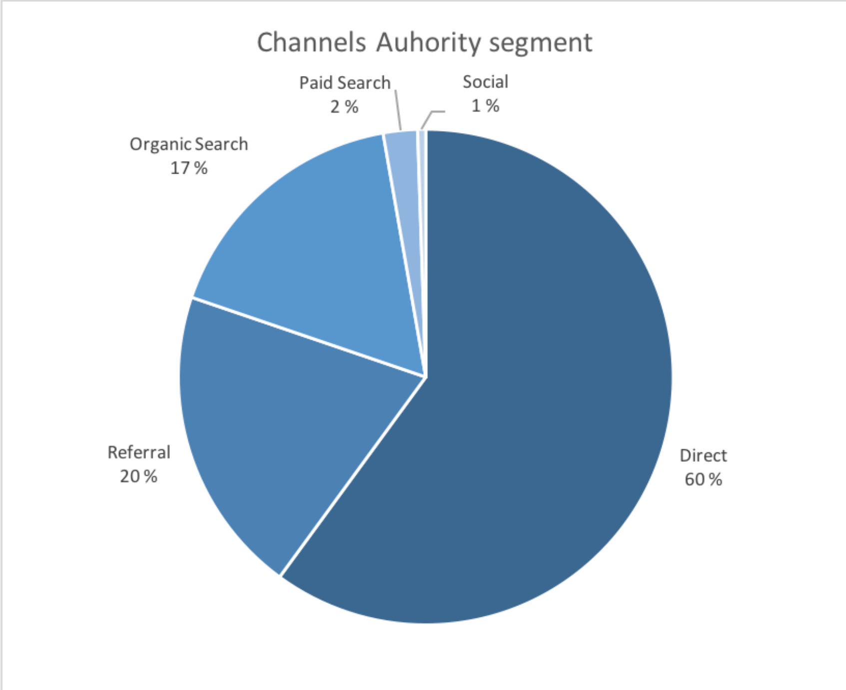


Figure 30: Channels from Authority segment

Inspecting the differences between Applicant and Authority segments, Authority segments' user are more inclined to the Direct channel (60%) compared to the Applicant segment (40%). 28 percent of the Applicant segment users arrive through Referral channel compared to the 20 percent of the users in the Authority segment. Interesting detail is that Social channel is present in the Authority segments channels, suggesting that users in this segment might be more actively following the service in various social media services such as Facebook or Twitter.

5.9. Engagement based on bounce rate

The bounce rate metric is a single-page session in which the users does not have further interactions with the website (Google, 2017a). High bounce rate might mean that the site is not providing essential content or functions for the users. In the case of Lupapiste, high bounce rate could mean that users are not able to achieve the desired goals.

The low levels of bounce rate throughout the timeline suggests that users are engaged with the service, shown in the Figure 31. The overall bounce rate for all users hovers above 20 percent line throughout the timeline (1st March 2015 to 31st March 2017). When applying segments to the bounce rate data, it is visible that the Applicant segment's bounce rate has been quite steadily around one percent between March 2015 and March 2017. On the other hand, the Authority segment's bounce rate has had more variance and showing an increase in the value from October 2016 onwards. Authority segment's bounce rate has risen and levelled around 8 percent in the past few months. It is difficult to extract explanations for the increased bounce rate from the analytics data, but it is more important to note the change in the bounce rate when developing the service in the future. Bounce rate could be used as a metric to evaluate the successfulness of new features or improvements made to the service.

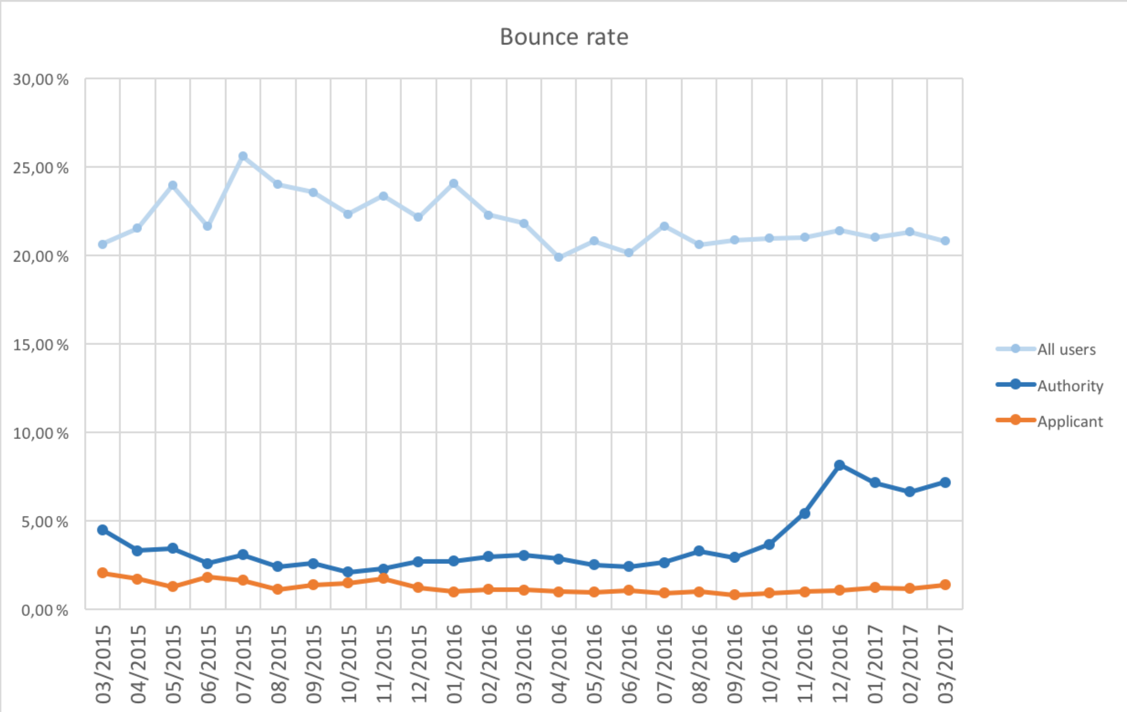


Figure 31: Bounce rate per segment

5.10. Engagement based on visit duration

Evaluating the Lupapiste service from the view point of user engagement, the average session duration data was exported from the Google analytics and transformed in to a graph shown in Figure 32.

The sessions includes all user interactions before thirty minutes of inactivity in Google analytics (Google, 2017b). In Figure 32, the average session durations of Authority and Applicant segments are presented revealing big differences between the two aforementioned user segments. On average, sessions in authority segment are 45% longer compared to the Applicant user segment. With the Authority segment, the average session duration is 20 minutes 37 seconds while Applicant segments' average session duration is merely 9 minutes 42 seconds.

The Applicant segment's session duration across the timeframe (1st March 2015 – 31th March 2017) has been quite stable following the all-time average session duration (09:42). On the other hand, the Authority segment has more variety across the timeframe visible in Figure 32. Starting from January 2016, there is a climb in the sessions duration until the July 2016. Same kind increase is also present starting from January 2017 forward. The increasing session duration might be due to the fact that the constructing itself is planned to begin on the spring time and therefore Authorities are receiving and handling the applications before the spring time.

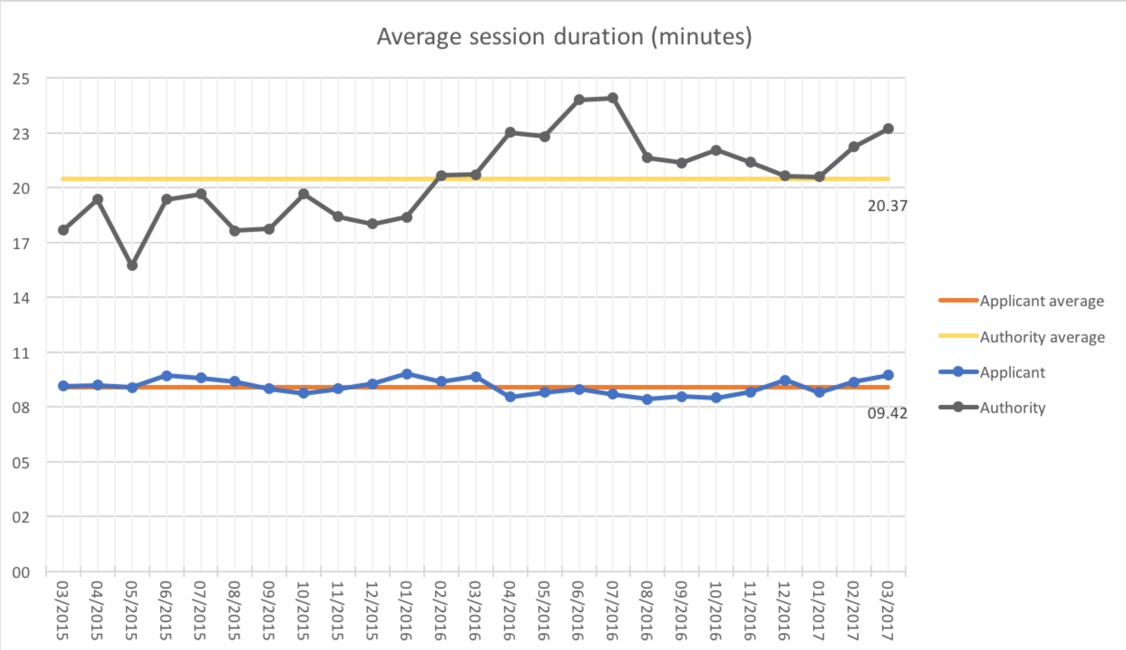


Figure 32: Average session duration

As the average session length is long, especially with the Authority segment's users, this warrants for the inspection of the page load times as users might be accessing multiple pages while using the Lupapiste service. Throughout the timeframe, the average pages per session is shown in Table 6 below. Authority segments number of pages per session (15.69) as well as the Application segments number of pages per session (9.27) confirms that the page load times could hinder the user experience if the load times are long.

Authority segment	Applicant segment
15.69	9.27

Table 6: Pages per session

The following Figure 33, inspect the average page load times for the Authority and Applicant segments. As seen from the Figure 33, there has been development in the average load time throughout the time frame. Average page load time for the Authority segment is 3,62 seconds and 2,10 seconds for the Applicant segment. More interestingly trend of the load time is decreasing, but includes peaks especially with the Authority segment. Starting from April 2016 until November 2017 higher than the average load time. Inspecting the values from the last three months, the average page load time has been decreasing.

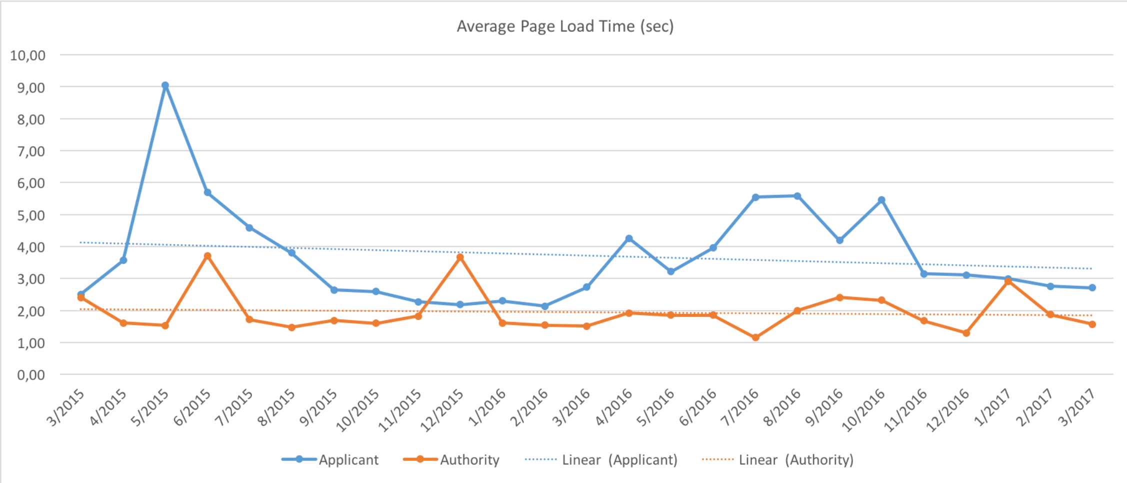


Figure 33: Average page load times

The average page load time is important metric to be followed actively as the longer load time can affect the user experience negatively. From the users point of view, waiting for a service to load content eventually will frustrate the users. In the case of Lupapiste optimizing heavily to decrease the page load time is important as users might not have alternative solution for the applying the construction permit. As for the Authority segment, the page load times can be seen even more critical, as Authority users are using the service as a professional tool. In the future, the page load times should be reported on daily basis in order to actively optimize the service if the metric is moving to an undesired direction.

6. Discussion

This research involved the use of web analytics data in order to understand how users are interaction with the Lupapiste service. Lupapiste web service enables municipalities and cities to provide an electronic service for the applying for a construction permit. Citizens are faced with a single point of contact for the construction permit process, enabling transparent and interactive process throughout the lifecycle of the construction application. Cities and municipalities are able to handle the application and the communications with the applicants altogether in the Lupapiste service. Also, handling and archiving different attachments and documents relating to the applications can be done through the Lupapiste service.

To evaluate the user experience and user engagement of the Lupapiste service, web analytics software was used to capture the user interactions with the service. Web analytics data were gathered, transformed and analyzed in order to understand the user behavior. The aim of the analysis was to create insights of the user behavior, which could be used, for example, to improve the service.

The case-study performed in this thesis was open-ended and explorative in which the research questions formed the basis for the research. The research was not solely restricted by the research question and the scope of the research was expanded when interesting phenomenon were discovered from the web analytics data. Much of the research relied on the Google Analytics and the standard and custom reports provided by Google Analytics. The Google Analytics reports were the main data source used in the research. Further transformations and visualizations were made with a spreadsheet software for more detailed information about the user behavior of the Lupapiste service. The starting point for the research was to inspect what can be learned about the users in the eGoverment context by examining the basic metrics in the web analytics software used for the research.

Inspecting any given website utilizing web analytics software, such as Google Analytics, there are two main characterizations of the data, namely sessions and users. A session contains all interactions that take place before 30 minutes of inactivity. From the user's perspective, the interactions might take place in one continuous time frame but includes a 30 minutes' break. As the session expires after the 30 minutes of inactivity, a new session will be initiated and counted. The user data, on the other hand, enables to group and link the individual sessions into an anonymous user via a generated identifier which is generated on the first visit and saved in to a cookie. Whenever a previously identified users return to the service, later sessions and interactions can be grouped to the same users.

When inspecting data with the session metric and seeing an increase, it is possible to conclude that the service is merely generating new sessions. The increase, on the other hand, can be seen as natural progression as the Lupapiste service's usage is opened in new cities and municipalities. The trend in the session amount can be seen as more important insight. The fact that the trend is upward, is an important indication of the successfulness of the expansion of the Lupapiste service. Also, the opposite downward trend should raise questions in the sense has the service already reached the peak of the sessions.

The amount of users on the other hand enables to inspect the web analytics data from a more concrete level. The user level enables to inspect the data in more relevant way, depicting a real situation of a user arriving to the service and interacting with the system. As was discussed in chapter 5.1, the difference between the user amount and session amount is quite significant, then how to interpret the difference in a meaningful way? Simply using the user amount as a reported metric might be good practice because the number is more concrete as users are using the service and not the sessions that the users generate.

Since the users and sessions metrics are general level metrics, in order to discover possible differences in the user behavior, the segmentation of the web analytics data enables the inspection of the interactions relating to the service. In the case of the Lupapiste, the users can be divided into two distinctive segments. The segments represent two real user groups each with its own needs, goals and motivations concerning Lupapiste. In order to discover the possible differences in the usage of the service and the user interaction, the segments were utilized in the extraction and comparison of the web analytics data. The usage times during the week days and hours of the days revealed interesting differences between the two segments. Although, the usage times also had similarities between the segments.

Event tracking enables one to inspect the user behavior in a detailed fashion. What specific interactions takes place in the service and what functions and features users are actually using in the service. Event categories allows one to give context for the user interactions. By examining the event categories with the combination of event actions, it is possible to evaluate where and what functions users are using. Also, events actions and even labels enable far more detailed inspection of the interactions. Event actions combined with event labels essentially answer the question what users do while interacting the service.

When considering the future improvements and the development of the Lupapiste service, the event data can provide a guideline in the sense that what sections of the service are used the most. Thus, enhancing the most used features and sections would benefit the users the most, based on the real usage. Furthermore, even data can be the basis of designing future usability studies, while fixing the issues found from the studies, would enhance user experience the most from the users' point of view. On the other hand, the findings from a usability studies, when performed with a small number of participants, further validations of the findings could be performed by utilizing the web analytics data. As the web analytics data can be captured essentially from the whole user base, the sample size can be unlimited.

Examining only the number of events does not provide information about the user behavior in the sense in which order the events takes place. By examining the flow of events in the service, more insightful information about the user behavior can be extracted. In this thesis, findings suggest that the user behavior in the service resembles the so called pogo-sticking behavior in which the usage is not following any linear process. The reasons for the pogo-sticking behavior, as to why users behave this way, would need attitudinal data gathering, for example, surveys, interviews or contextual observations.

As mobile phones have become an alternative device for browsing the web in addition to desktop computers, the use of small screen devices might require additional design decisions to be made. By inspecting the shares of different device categories and the details of each device category, what is visible is that the small screen devices are used only for shorter sessions. What this might suggest is that users are coming to the service to take a peek at the current status or they have received notifications requiring actions from the users. However, the details of the desktop device category suggest that the actions are done with a desktop device. In the future, the share of mobile devices will potentially increase based on the amount of sold devices. It should be considered in the future whether to optimize the whole service for mobile devices or to support the current behavior of previewing the current status of the application and the service.

Users can arrive to the Lupapiste through various channels. Some of the cities and municipalities promote the service on their websites in order to guide the users to the service. Still, users can arrive to the service for example by using common search engines. Although, the URL of the service could be used directly accessing the service. Segmented web referral data from the web analytics revealed interesting differences between the two segments. Applicant users seem to be more incline to utilize the search engines and links on the websites of the municipalities and cities. However, the Authority segment users arrive directly to the service either using the URL or bookmark from the web browser.

The measurement of the user engagement of Lupapiste indicates that Authority users are using the service on daily basis for extensive time period. Suggested by the bounce rate and average session length, users arrive to the service to advance their construction application given the state that the application currently is in. Nonetheless, the Applicant users are also engaged with the service. Since the usage times of Lupapiste is quite long with both of the segments, it is important to follow-through the load times of the service in order to create fluent experience for the users.

The evaluation of the current data collection implementation concerned inspecting the number of pages, event categories and event actions currently begin tracked by the web analytics software. Based on the pages and events being tracked, the coverage of the web analytics implementation was not straightforward to evaluate. The documentation of the web analytics implementation would have enabled more exhaustive evaluation of the implementation, for example, whether some views of the service are missing the event tracking or the page tagging snippet altogether. The web analytics implementation should be documented extensively which would enable to evaluate and further develop the implementation of web analytics. When new views or pages are added to the service, these pages and views should be listed in documentation in combination with the possible event categories, event actions and event labels. Web analytics documentation would also enable transparency within the development team concerning the web analytics implementation.

Although, web analytics data can create insights about the user behavior, the usefulness of web analytics in context of construction permit application became more evident when the web analytics data was inspected based on the more detailed event tracking data. When the contexts of the events were examined, more detailed information about the service were discovered. By utilizing event data in the future, the Lupapiste service could be improved by enhancing the views and functions which are used the most by the users. Furthermore, segmenting the event data revealed differences between the two segments. By evaluating the segmented data, the future improvements can be targeted for a specific segment. Also, the segmented event data can be utilized when considering which of the segments could gain the most value out of the planned improvements.

7. Conclusions

In order to create a realistic scenario of the user behavior of any given web site, product or application, web analytics can be utilized to automatically capture the user interactions. Although, utilization of web analytics might be more familiar in the eCommerce context, the use of web analytics can provide useful information and insights from the real usage of the service in any given context.

The use of web analytics can start by simply posing a question about the user behavior. The question then guides the data gathering and transforming the data to the required format followed by the analysis of the data to answer the question.

Modern web analytics tools allow the inspection of the behavioral data from different viewpoints by enabling temporal filters to the data. With the use metrics, dimensions and reports, it is possible to create a detailed view of the user behavior.

Web analytics data can be utilized in order widen the knowledge of the web site or web service being inspected, essentially learning how the users are actually using the service. Furthermore, the web analytics data and insights should be used as a combination: web analytics data can guide the further research effort by creating a solid base for usability studies. Also, the findings from the usability studies can be validated through the web analytics data. Still the web analytics data cannot provide information about the motivations of the users. Web analytics data should be combined with attitudinal data in order to learn *why* user behave the way they do.

By examining the web analytics data from different viewpoints, the details of the user behavior can be revealed and transformed into insights for considering the future improvements of the service. Segmenting the web analytics also reveals the possible details concerning the different users of the service.

The research in this thesis inspected user interaction data captured from the Lupapiste web service. The web analytics data was analyzed from various viewpoints, for example, the number of sessions and users was inspected in order to see changes over a timeframe. By segmenting the web analytics data, it was possible to inspect the differences between the different user groups of the service. Utilizing the event tracking data of the Lupapiste service, enabled to view the user behavior and the user interactions with the service in a more detailed level. The importance of the documentation of the web analytics was also raised when the coverage and thoroughness of the automatic data collection implementation was evaluated.

From the user experience designer's view point the web analytics data is a concrete tool presenting the details of the user behavior. The details can be transformed in to design decisions for the future, by examining what sections or features users are using, thus improving the most used features would enhance the overall user experience the most.

As noted by Beasley (2013) web analytics should be one of the central skills user experience designers working with websites, web services and mobile applications should master (Beasley, 2013, p. 227). In the modern world of both web and mobile development the constant change should be embraced by utilizing new tools and methods in order to learn more about the user behavior.

Web analytics data enables to inspect the user behavior from different level of detail. By using different time frames, the data reveals possible changes in the behavior. Furthermore, when the data is viewed from the event context, the intricacies of the user behavior is revealed. When applying improvements to the service, web analytics is an efficient way for evaluating the change over time and target the improvements to the most valuable features.

References

Arshad, A., & Ameen, K. (2015). Usage patterns of Punjab university library website: A transactional log analysis study. *The Electronic Library*, *33*(1), 65–74. https://doi.org/10.1108/el-12-2012-0161

Association, W. A. (2008). Web Analytics Definition. Retrieved from http://www.digitalanalyticsassociation.org/Files/PDF_standards/WebAnalyticsDefinitions.pdf

Attfield, S., & Kazai, G. (2011). Towards a science of user engagement (Position Paper). *WSDM Workshop on User Modelling for Web Applications*.

Beasley, M. (2013). *Practical Web Analytics for User Experience*. *Practical Web Analytics for User Experience*. Elsevier. https://doi.org/10.1016/B978-0-12-404619-1.00004-6

Cardello, J. (2013). Three Uses for Analytics in User-Experience Practice. Retrieved May 15, 2017, from https://www.nngroup.com/articles/analytics-user-experience/

Castellano, G., Fanelli, A., & Torsello, M. (2013). Web Usage Mining: Discovering Usage Patterns for Web Applications. *Advanced Techniques in Web …*, 75–104.

Chen, A. (2012). Know the difference between data-informed and versus data-driven. Retrieved April 17, 2017, from http://andrewchen.co/know-the-difference-between-data-informed-and-versus-data-driven/

Churchill, E. F. (2012). From Data Divination to Data-Aware Design. *Interactions*, *19*(5), 10. https://doi.org/10.1145/2334184.2334188

Croll, A., & Yoskovitz, B. (2013). *Lean Analytics: Use Data to Build a Better Startup Faster*. O'Reilly Media, Inc.

Das, R., & Turkoglu, I. (2009). Creating meaningful data from web logs for improving the impressiveness of a website by using path analysis method. *Expert Systems with Applications*, *36*(3 PART 2), 6635–6644. https://doi.org/10.1016/j.eswa.2008.08.067

Evolta Oy. (2017a). Lupapiste. Retrieved May 15, 2017, from https://www.lupapiste.fi/

Evolta Oy. (2017b). Services. Retrieved May 15, 2015, from http://evolta.fi/en/solution/

Farney, T. (2016). Using Google Tag Manager in Your Library. *Library Technology Reports*, *52*(7), 2,9-13.

Godin, S. (2014). Analytics without action. Retrieved October 22, 2016, from http://sethgodin.typepad.com/seths_blog/2014/08/analytics-without-action.html

Google. (2016). Adding analytics.js to Your Site. Retrieved March 15, 2017, from https://developers.google.com/analytics/devguides/collection/analyticsjs/

Google. (2016). Event Tracking. Retrieved May 15, 2017, from https://developers.google.com/analytics/devguides/collection/analyticsjs/events

Google. (2016). The Javascript tracking snippet. Retrieved May 1, 2017, from https://developers.google.com/analytics/devguides/collection/analyticsjs/#the_javascript_tracking_snippet

Google. (2017a). Bounce rate. Retrieved May 20, 2017, from https://support.google.com/analytics/answer/1009409?hl=en

Google. (2017b). How a web session is defined in Analytics. Retrieved May 15, 2017, from https://support.google.com/analytics/answer/2731565?hl=en

Google. (2017c). Tag Manager Help. Retrieved April 11, 2017, from https://support.google.com/tagmanager/answer/6106924?hl=en

Hassenzahl, M. (2010). *Experience Design: Technology for All the Right Reasons*. *Synthesis Lectures on Human-Centered Informatics* (Vol. 3). https://doi.org/10.2200/S00261ED1V01Y201003HCI008

Hern, A. (2014). Why Google has 200m reasons to put engineers over designers. Retrieved May 15, 2017, from https://www.theguardian.com/technology/2014/feb/05/why-google-engineers-designers

Hoverfält, O. (2015). Data-Driven Design Day 2015: Olof Hoverfält - Building a data-driven culture. Retrieved April 17, 2017, from https://www.youtube.com/watch?v=cmUmk4z63L8

Hurst, M. (2010). Search ROI's missing element: Search analytics. *Journal of Digital Asset Management*, *6*, 327–331. https://doi.org/10.1057/dam.2010.43

Jansen, B. J. (2009). *Understanding User –Web Interactions via Web Analytics*. *Synthesis Lectures on Information Concepts, Retrieval, and Services #6*. Morgan & Claypool Publishers.

Kumar, A. (2010). Yahoo! Digits: A Design Driven to Provide Instant "Data Driven Insights" and Its Use in User Experience Design. In *Proceedings of the 2010 International Conference on Interaction Design & International Development* (pp. 108–113). Swindon, UK: BCS Learning & Development Ltd.

Lalmas, M., O'Brien, H., & Yom-Tov, E. (2014). *Measuring User Engagement. Synthesis Lectures on Information Concepts, Retrieval, and Services* (Vol. 6). Morgan & Claypool Publishers. https://doi.org/10.2200/S00605ED1V01Y201410ICR038

Lazar, J., Feng, J. H., & Hochheiser, H. (2009). *Research Methods in Human-Computer Interaction, eTextbook*. Chichester, UNITED STATES: Wiley Textbooks.

Lehmann, J., Lalmas, M., Yom-Tov, E., & Dupret, G. (2012). Models of user engagement. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7379 LNCS*, 164–175. https://doi.org/10.1007/978-3-642-31454-4_14

MarketsAndMarkets. (2016). Web Analytics Market worth $3.09 Billion by 2019. Retrieved May 30, 2017, from http://www.marketsandmarkets.com/PressReleases/web-analytics.asp

Nakatani, K., & Chuang, T. (2011). A web analytics tool selection method: an analytical hierarchy process approach. *Internet Research*, *21*(2), 171–186. https://doi.org/10.1108/10662241111123757

Pavliscak, P. (2014). Data-Informed Design: Three Data Stories. Retrieved April 15, 2017, from http://www.uxmatters.com/mt/archives/2014/11/data-informed-design-three-data-stories.php

Pavliscak, P. (2015). *Data-Informed Product Design*. O'Reilly Media, Inc. https://doi.org/978-1-491-93129-5

Román, P. E., Dell, R. F., Velásquez, J. D., & Loyola, P. S. (2014). Identifying user sessions from web server logs with integer programming. *Intelligent Data Analysis*, *18*(1), 43–61. https://doi.org/10.3233/IDA-130627

Singal, H., Kohli, S., & Sharma, A. K. (2014). Web analytics: State-of-art & literature assessment. *2014 5th International Conference - Confluence The Next Generation Information Technology Summit (Confluence)*, 24–29. https://doi.org/10.1109/CONFLUENCE.2014.6949041

StatCounter. (2017). Desktop vs Mobile vs Tablet Market Share in Finland. Retrieved May 20, 2017, from http://gs.statcounter.com/platform-market-share/desktop-mobile-tablet/finland

Vellingiri, J., Kaliraj, S., Satheeshkumar, S., & Parthiban, T. (2015). A Novel Approach for User Navigation Pattern Discovery and Analysis for Web Usage Mining. *Journal of Computer Science*, *11*(2), 372–382. https://doi.org/10.3844/jcssp.2015.372.382

W3Techs. (2015). Usage of traffic analysis tools for websites. Retrieved May 18, 2015, from http://w3techs.com/technologies/overview/traffic_analysis/all

Waisberg, D., & Kaushik, A. (2009). Web Analytics 2.0: empowering customer centricity. *The Original Search Engine Marketing …*, *2*(1), 7.

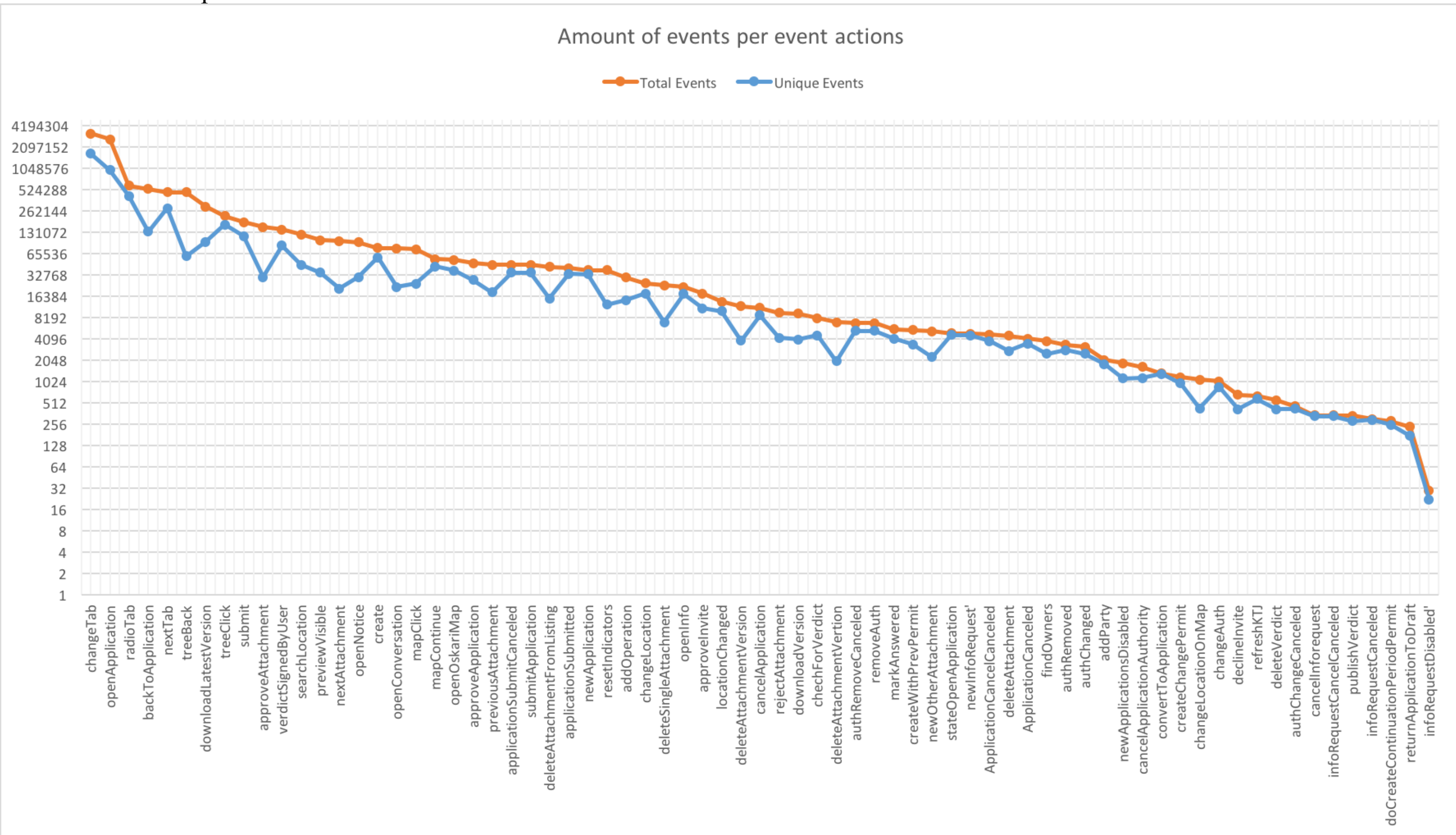
Walker, A. (2009). Google's Marissa Mayer Assaults Designers With Data. Retrieved May 15, 2017, from https://www.fastcompany.com/1403230/googles-marissa-mayer-assaults-designers-data

Warnock, D., & Lalmas, M. (2015). An Exploration of Cursor tracking Data, 1–31.

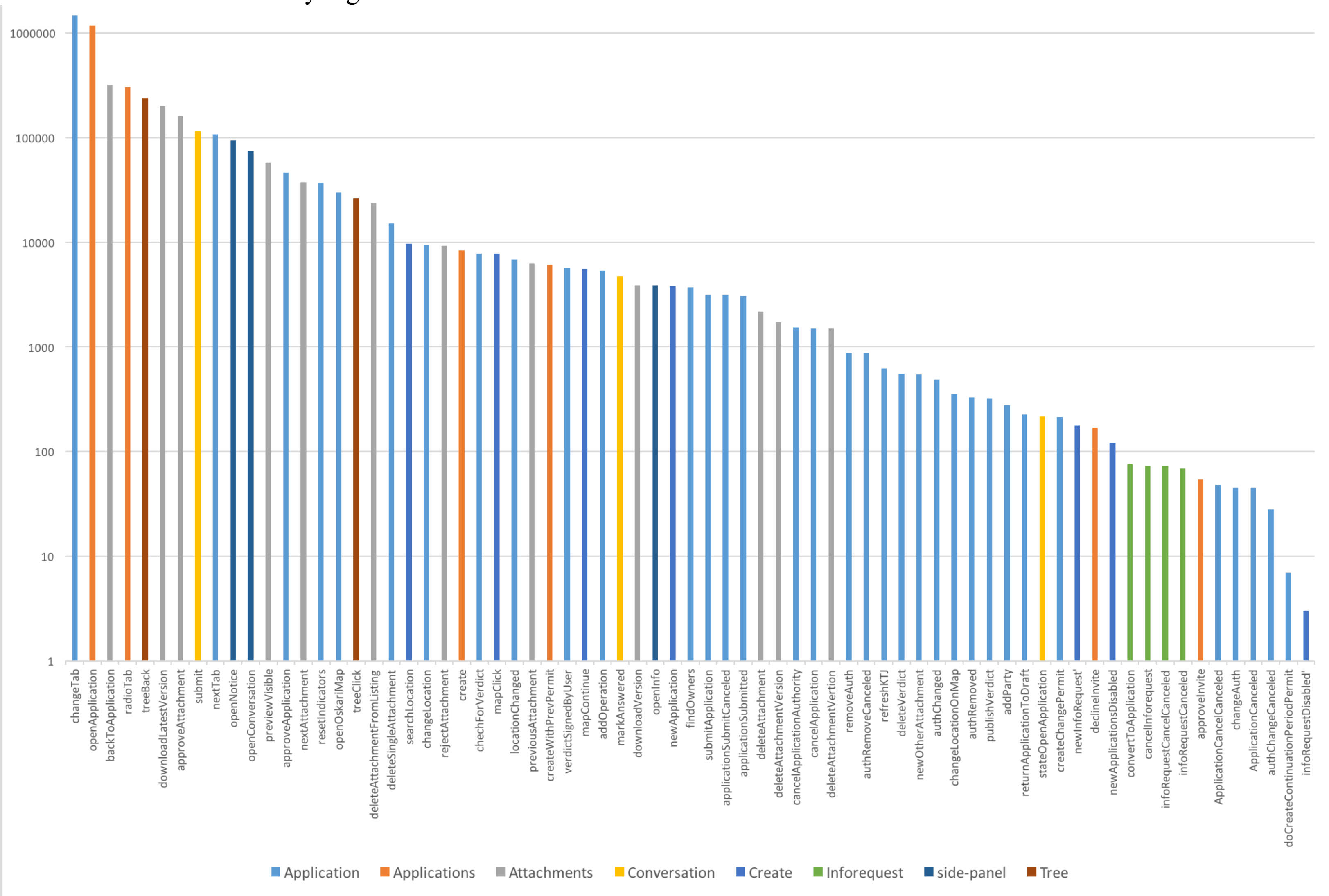
Wikipedia. (2017). E-government.

Appendix 1

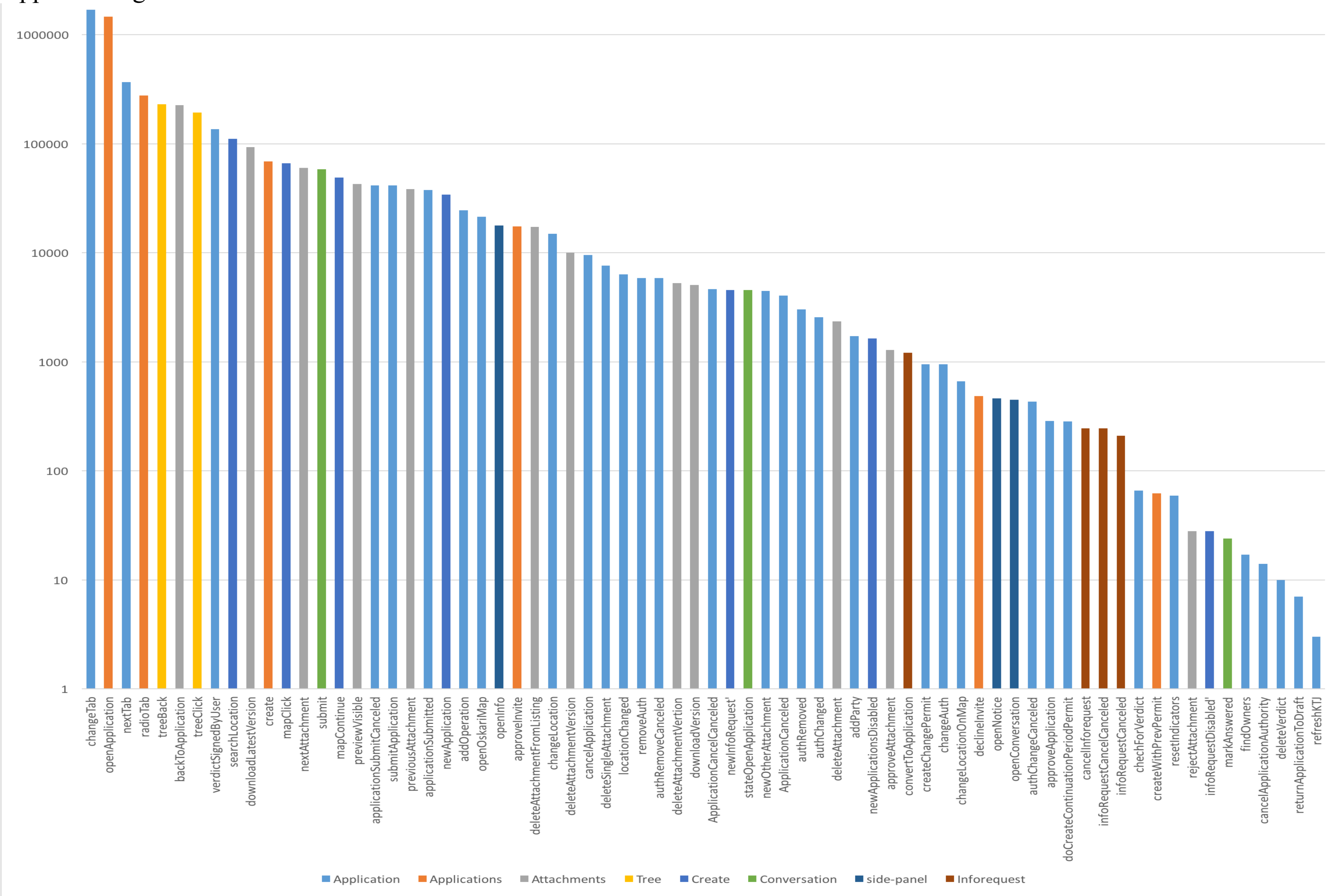
Number of events per event actions



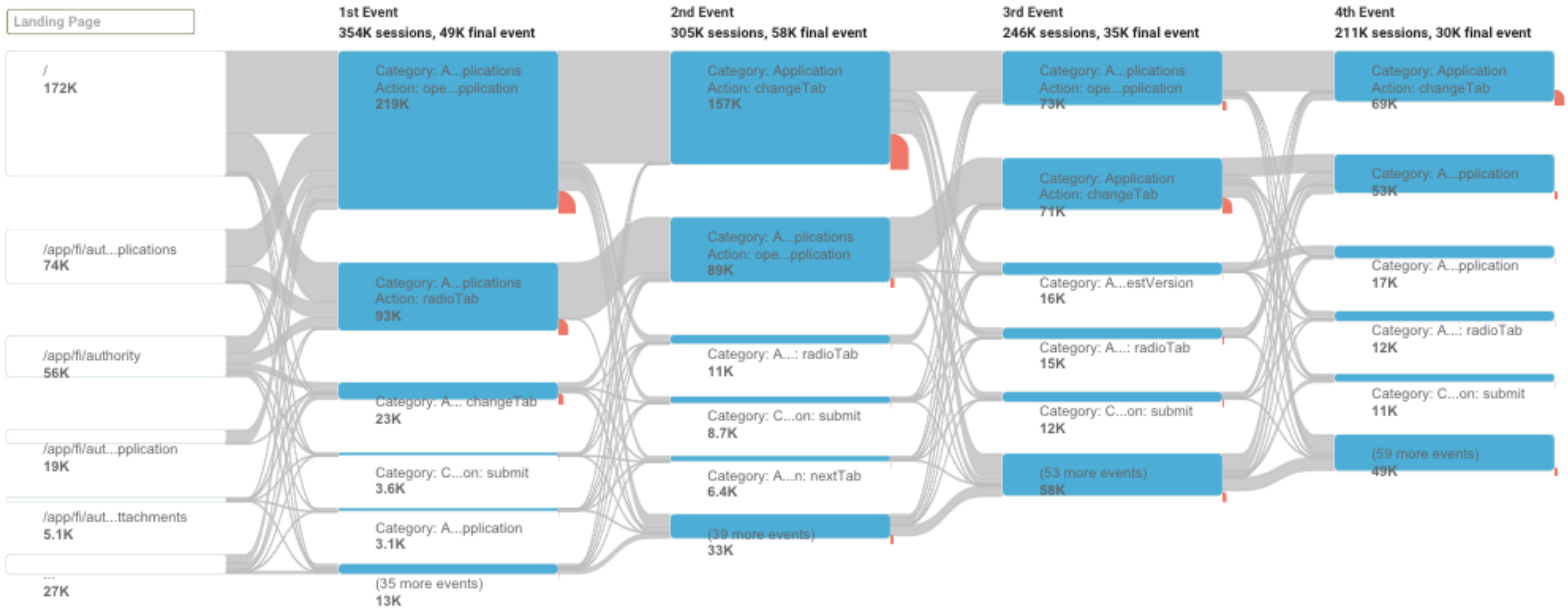
Event actions of the Authority segment



Applicant segment’s event actions



Event flow of the Authority segment



Event flow of the Applicant segment